Level set Cox processes

Anders Hildeman[1], David Bolin[1], Jonas Wallin[2], and Janine B. Illian[3]

[1]Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Sweden
[2]Department of Statistics, Lund University, Sweden
[3]School of Mathematics and Statistics, University of St Andrews, Scotland, UK

Abstract

An extension of the popular log-Gaussian Cox process (LGCP) model for spatial point patterns is proposed for data exhibiting fundamentally different behaviors in different subregions of the spatial domain. The aim of the analyst might be either to identify and classify these regions, to perform kriging, or to derive some properties of the parameters driving the random field in one or several of the subregions. The extension is based on replacing the latent Gaussian random field in the LGCP by a latent spatial mixture model specified using a categorically valued random field specified through level set operations on a Gaussian random field. This allows for standard stationary covariance structures, such as the Matérn family, to be used to model random fields with some degree of general smoothness but also occasional and structured sharp discontinuities.

A computationally efficient MCMC method is proposed for Bayesian inference and we show consistency of finite dimensional approximations of the model. Finally, the model is fitted to point pattern data derived from a tropical rainforest on Barro Colorado island, Panama. We show that the proposed model is able to capture behavior for which inference based on the standard LGCP is biased.

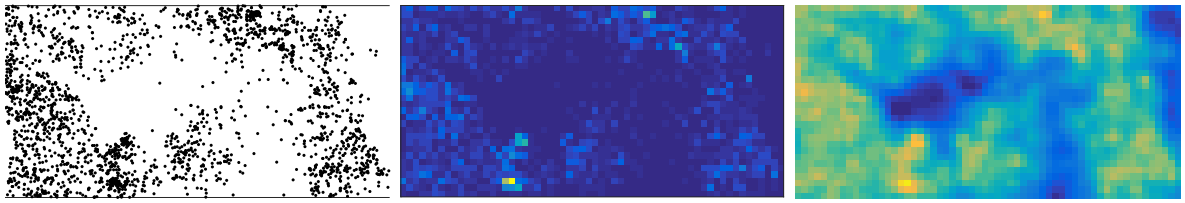
1 Introduction

Cox processes, and in particular log-Gaussian Cox processes (LGCP), have been used extensively as flexible models of spatial point pattern data [39, 38, 31, 20]. These are hierachical point process models where the point locations are assumed to be independent given a random intensity function

$$\lambda(\boldsymbol{s}) = \exp\{\mathrm{B}(\boldsymbol{s})\,\beta + \mathrm{X}(\boldsymbol{s})\}, \tag{1}$$

where $\mathrm{B}(\boldsymbol{s})$ is a, possibly multivariate, function of covariates and $\mathrm{X}(\boldsymbol{s})$ is a Gaussian random field, which is typically assumed to be stationary. The random field captures spatial structure in the point pattern that the given covariates cannot capture. In this work, we relax the assumption that a single stationary Gaussian field can account for those remaining spatial structures and develop a mixture model based on level set inversion.

To motivate the relevance of the approach we consider a point pattern data set formed by the locations of the tree species *Beilschmiedia Pendula*, one of the species in the tropical rainforest plot on Barro Colorado Island [12, 14, 10, 28]. The point pattern comprises 2461 point locations in a rectangular observation window (500 m $\times$ 1000 m), see Figure 1a. This pattern has been analysed repeatedly in the literature and is one of the example patterns in the `R` [1] package *spatstat* [5]. Previous analyses have fitted a log-Gaussian Cox process [38] to this and related



(a) (b) (c)

Figure 1: Spatial point pattern formed by the locations of trees of the species *Beilschmiedia pendula* in a 500 m $\times$ 1000 m rainforest plot on Barro Colorado Island (a), a gridded version of the data (b), and posterior mean of log intensity using a log-Gaussian Cox process model (c).

data sets to draw conclusions on the association of habitat preferences based on a number of spatial covariates reflecting local soil chemistry and topography [38, 31].

On close inspection, the pattern shows large areas of very low point intensity where hardly any trees can be found. Anecdotal knowledge reveals that these regions are covered by a swamp, where the tree species is known to be very unlikely to grow, independent of local soil covariates and topography. However, data on the exact extent of the swamp is not available. We initially fitted a log Gaussian Cox process to the data, with an intensity function as in (1) using 11 covariates, and the estimated posterior mean of the model also predicts large regions of low intensity, as plotted in Figure 1c. However, when a LGCP model that ignores the presence of swamp is fitted to this pattern, the swamp is likely to act as a confounding factor and this is likely to impact on inference. Hence, any conclusions on habitat preferences of the species will be heavily biased. Covariates associated with the presence of the swamp may appear to have a significant correlation with the intensity of the tree growth, or important covariates might appear non-significant as they vary indepedently of the presence of the swamp.

The approach we take here is designed to capture sharp discontinuities in the intensity surface that result from qualitative yet unavailable covariates or environmental conditions as the one seen in this example. Further examples where this approach could be important is ecological data with several distinct types of habitat, spatial regions with different treatment regimes in medical data, or finding regions of interest in biology [23]. Specifically, we consider a Cox process model where the intensity surface is modeled using a Bayesian level set approach. The proposed model is an extension of the log-Gaussian Cox process with increased flexibility resulting from a random segmentation of the spatial region into K classes. The intensity surfaces of the regions associated with the K different classes can be modeled separately of each other by log-Gaussian random fields with simple covariance structures, while still maintaining flexibility. We refer to the proposed model as the *level set Cox process (LSCP)*.

Finding sharp discontinuities in the intensity surface is similar to the level set inversion problem [49, 9] in the inverse problems literature, where the main objective is to find interfaces between geometrical regions based on observed data. Level set inversion has been used extensively for segmentation [11, 36, 50], for multiphase flow modeling [7, 19], and for statistical modeling of porous materials [40]. The interfaces in the level set inversion approach are modeled as level sets of an unknown *level set function*. Higgs and Hoeting [24] modeled spatially correlated categorical data using a Bayesian level set approach, where the level set function was modeled as a Gaussian random field. This probabilistic approach, which Iglesias et al. [29] and Dunlop et al. [21] extended to more general inverse problems, has the advantage that the level sets can be estimated through the posterior distribution of the level set function given the observed data. The LSCP model could be viewed as an extension of these approaches to point pattern data.

The LSCP is, like the LGCP, a process defined on a continuous domain. In order to use the model in practical inference some finite dimensional approximations are required. We show that the classical lattice approximation that is often used for LGCP models converges, in total variation distance, to the continuous model as the grid gets finer also for the LSCP models. We further propose a computationally efficient Markov chain Monte-Carlo (MCMC) algorithm for Bayesian inference on the model parameters, based on preconditioned Crank-Nicholson Langevin proposals [16].

This paper is structured as follows. A detailed model description is given in Section 2. In Section 3, we derive the MCMC algorithm for the Bayesian inference. Section 4 analyses the *Beilschmiedia Pendula* point pattern of rainforest trees with the new approach. Finally, Section 5 discusses the presented material and possible future extensions of it. The theoretical results and proofs are given in two appendices.

2 The LSCP model and its properties

In this section, we first introduce the LSCP model in Section 2.1. Some examples of the model are presented in Section 2.2 and basic properties of the model are presented in Section 2.3. Finally, Section 2.4 introduces finite dimensional approximations necessary for inference.

2.1 Model definition

The Bayesian level set inversion problem of Iglesias et al. [29] corresponds to reconstructing a latent field of the form

$$\mathrm{X}(\boldsymbol{s}) = \sum_{k=1}^{K} \mathrm{X}_k \, \mathbb{I}\left(\boldsymbol{s} \in \mathcal{E}_k\right), \tag{2}$$

on a bounded domain $\mathcal{D} \subset \mathbb{R}^2$ given noisy data. Here $\mathcal{E}_k \subset \mathcal{D}$ is the spatial region associated with segmentation class k, and X_k are fixed values. If the constants $\{\mathrm{X}_k\}_{k=1}^K$ are known, the partition $\{\mathcal{E}_k\}_{k=1}^K$ characterizes X. Iglesias et al. [29] defined $\mathcal{E}_k$ as an excursion set of an unknown random continuous *level set function*, X_0, such as $\mathcal{E}_k = \{\boldsymbol{s} : \mathrm{c}_{k-1} < \mathrm{X}_0(\boldsymbol{s}) \leq \mathrm{c}_k\}$. Here c_k are constants such that $\{-\infty = \mathrm{c}_0 < \mathrm{c}_1 < \ldots < \mathrm{c}_{K+1} = \infty\}$ and X_0 is assumed to be a realization of a Gaussian random field. Thus, this model corresponds to the level set problem for categorical data by Higgs and Hoeting [24]. The level set model using a latent Gaussian random field is not identifiable with regards to the parameter triplet *threshold values*, *mean*, and *marginal variance* of the level set field, X_0. Hence, we define X_0 to have standard normal marginal distributions in order to make the model identifiable.

We extend the level set function of (2) by replacing the fixed constants X_k by Gaussian random fields and denote these by $\mathrm{X}_k(\boldsymbol{s}) + \mu_k(\boldsymbol{s})$, where μ_k is a deterministic mean function and X_k is a centered Gaussian random field with covariance function $\mathcal{C}_k$.

$$\mathrm{X}(\boldsymbol{s}) = \sum_{k=1}^{K} \left(\mathrm{X}_k(\boldsymbol{s}) + \mu_k(\boldsymbol{s})\right) \mathbb{I}\left(c_{k-1} < X_0(\boldsymbol{s}) + \mu_0(\boldsymbol{s}) \leq \mathrm{c}_k\right). \tag{3}$$

This can be regarded as a mixture model of Gaussian fields related to the non-stationary geostatistical model proposed by Fuentes [22]. We use this model to specify a statistical model for spatial point process data through a Cox process [20], modeling the number of occurrences of some event in a subregion $\mathcal{E} \subseteq \mathcal{D}$ as an inhomogeneous Poisson process conditional on a

realization of X, i.e.

$$\mathrm{Y}(\mathcal{E}) \sim \mathsf{Pois}\left(\int_{\mathcal{E}} \lambda(\boldsymbol{s}) d\,\boldsymbol{s}\right),$$

where the intensity surface is $\lambda(\boldsymbol{s}) = \exp\{\mathrm{X}(\boldsymbol{s})\}$. For ease of notation we define $\mathrm{Z}_k(\boldsymbol{s}) = \mathbb{I}\left(\mathrm{c}_{k-1} < \mathrm{X}_0(\boldsymbol{s}) + \mu_0(\boldsymbol{s}) \leq \mathrm{c}_k\right)$.

A common usage of point process models is to study the effect of covariates on observed point patterns. A simple way of doing this is through a standard Poisson regression, where the log-intensity of the point process is of the form $\log \lambda(\boldsymbol{s}) = \mathrm{B}(\boldsymbol{s})\boldsymbol{\beta}$, where $\mathrm{B}(\boldsymbol{s})$ are the covariates of interest. This can easily be incorporated into the LSCP model by letting $\mu_k(\boldsymbol{s}) = \mathrm{B}(\boldsymbol{s})\boldsymbol{\beta_k}$ or $\mu_k(\boldsymbol{s}) = \mu(\boldsymbol{s}) = \mathrm{B}(\boldsymbol{s})\boldsymbol{\beta}$. To avoid identifiability issues with the threshold parameters, one should not use an intercept term in the fixed effects for the mean of X_0.

2.2 Specific cases

Poisson regression and LGCPs are special cases of the LSCP model. For an illustration of the flexibility of the model, Figure 2 shows the log-intensity for four special cases simulated on the unit square. In these figures, all Gaussian random fields are assumed to have constant means μ and Matérn covariance functions [37],

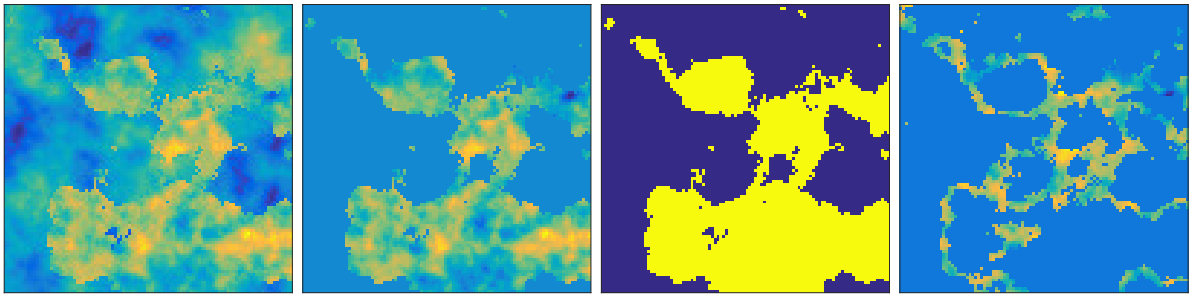
$$\mathcal{C}(h) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa h)^{\nu} K_{\nu}(\kappa h),$$

where $\sigma^2 = \mathrm{Var}(\mathrm{X}_k(\boldsymbol{s}))$, $\kappa = \frac{\sqrt{8\,\nu}}{\mathrm{r}}$, and ν is a smoothness parameter. Further, r is the correlation range approximately corresponding to the value of h where the correlation is 0.1, K_ν is a modified Bessel function of the second kind, and Γ is the Gamma function.

The patterns were generated using the same random seed such that the level set function is the same for all cases, yielding comparable results. A realization of $\log \lambda(\boldsymbol{s})$ using two classes is shown in panel (a). The log-intensity surface of the first class has $\mu = 2$ and $r = 0.1$, whereas the second class has $\mu = 0$ and $r = 0.2$. Both fields have $\sigma = \nu = 1$. The level set field, X_0, has a threshold value at the origin, $\mathrm{c}_1 = 0$, and range $\mathrm{r} = 0.4$. In the figure, the regions belonging to the two classes, and the difference in spatial correlation range is clearly visible.

A simplification of the model is obtained by assuming that the intensity for one of the two classes is constant (change $X_1(\boldsymbol{s})$ to a constant X_1, for instance). A realization of such a log-intensity surface can be seen in Panel (b). This model might be relevant in applications where some unknown factor makes it unlikely to observe points in certain subregions and may be regarded as spatially varying zero-inflation [33]. If a standard LGCP is fitted to data of this type some overdispersion will be unexplained and the estimated mean field and covariance parameters will be biased; this is not the case for the LSCP model. We discuss an example of this in Section 4. The two-class model can of course be simplified further by assuming a constant intensity for both classes, and $\log \lambda(\boldsymbol{s})$ is then of the form (2). This is identical to the model of [41] and corresponds to a special case of the *Random-set-generated Cox process* [30]. A realization of this model is shown in Panel (c).

The last model example uses the structure of the level set formulation to capture effects on the boundary between two regions. For a model with three classes, the second class takes on the role of an interface layer between the first and third class as can be seen in Panel (d). The log-intensity is in this case $\log \lambda(\boldsymbol{s}) = \sum_{i=1}^{3} \left(\mathrm{X}_i(\boldsymbol{s}) + \mu_i(\boldsymbol{s})\right) \mathrm{Z}_i(\boldsymbol{s})$. This can be used to model effects present on the boundary between two regions. Examples of potential applications are activity on shore lines between water and land or mixing regions between fluids.



(a) (b) (c) (d)

Figure 2: Realization of the log-intensity surface, $\log \lambda(\boldsymbol{s})$, for the four models presented in Section 2.2. Panel a) corresponds to the model with two random classes, Panel b) with one constant and one random, panel c) with two constant, and Panel d) is the model with two constant and a third random boundary class.

2.3 Model properties

The intensity measure $\Lambda = \{\Lambda(\mathcal{E}) = \int_{\mathcal{E}} \lambda(\boldsymbol{s}) d\,\boldsymbol{s}; \mathcal{E} \subseteq \mathcal{D}\}$ for a Cox process is well-defined if λ is almost surely finite and integrable. The LSCP model with $K = 1$ is just the standard LGCP model, which has a well-defined random intensity measure if realizations of the Gaussian field are identified with their continuous modifications [39]. For $K > 1$, a continuous modification does not need to exist but almost sure integrability follows if X_0 is a.s. continuous which ensures that the sets $\{\boldsymbol{s} : c_{k-1} < \mathrm{X}_0(\boldsymbol{s}) \leq \mathrm{c}_k\}$ are a.s. Lebesgue measurable for all $k \in \{1, ..., K\}$. Hence the LSCP model is well-defined when the realizations of all Gaussian fields are identified with their continuous modification with respect to the Lebesgue measure. By the same argument as in Theorem 3 of Møller et al. [39], ergodicity of the LSCP model follows from ergodicity of $\log \lambda$. Thus, the LSCP model is ergodic if all latent Gaussian fields are ergodic.

The following proposition gives semi-explicit formulas for the first and second order product densities.

Proposition 2.1. *For an LSCP with log-intensity* (3)*, where* $\{X_k\}_{k=1}^K$ *are zero-mean stationary random fields with covariance functions* $\{\mathcal{C}_k\}_{k=1}^K$*, the first moment of the intensity function equals*

$$\rho_1(\boldsymbol{s}) = \mathbb{E}\left[\lambda(\boldsymbol{s})\right] = \sum_{k=1}^{K} \exp\left(\mu_k(\boldsymbol{s}) + \frac{\mathcal{C}_k(0)}{2}\right) \left(\Phi\left(c_k - \mu_0(\boldsymbol{s})\right) - \Phi\left(c_{k-1} - \mu_0(\boldsymbol{s})\right)\right),$$

where Φ *is the CDF of a standard normal distribution. Further, the second moment of* λ*,* $\rho_2(\boldsymbol{s}_1, \boldsymbol{s}_2)$*, corresponding to the second order product density equals*

$$\begin{aligned}\rho_2(\boldsymbol{s}_1, \boldsymbol{s}_2) = &\sum_{k=1}^{K} \exp\left(\mu_k(\boldsymbol{s}_1) + \mu_k(\boldsymbol{s}_2) + \mathcal{C}_k(0) + \mathcal{C}_k(|\,\boldsymbol{s}_1 - \boldsymbol{s}_2\,|)\right) p_{kk} \\ &+ \sum_{k=1}^{K} \sum_{l \neq k} \exp\left(\mu_l(\boldsymbol{s}_1) + \mu_k(\boldsymbol{s}_2) + \frac{\mathcal{C}_k(0) + \mathcal{C}_l(0)}{2}\right) p_{lk}.\end{aligned}$$

Here

$$\begin{aligned}p_{lk} = &\Phi_d(\mathrm{c}_k - \mu_0(\boldsymbol{s}_1), \mathrm{c}_l - \mu_0(\boldsymbol{s}_2)) - \Phi_d(\mathrm{c}_{k-1} - \mu_0(\boldsymbol{s}_1), \mathrm{c}_l - \mu_0(\boldsymbol{s}_2)) \\ &- \Phi_d(\mathrm{c}_k - \mu_0(\boldsymbol{s}_1), \mathrm{c}_{l-1} - \mu_0(\boldsymbol{s}_2)) + \Phi_d(\mathrm{c}_{k-1} - \mu_0(\boldsymbol{s}_1), \mathrm{c}_{l-1} - \mu_0(\boldsymbol{s}_2)),\end{aligned}$$

where Φ_d is a CDF of a standard bivariate Gaussian distribution with correlation $d = \mathcal{C}_0(|\, \boldsymbol{s}_1 - \boldsymbol{s}_2 \,|)$.

The proof is given in Appendix B. Both Φ and Φ_d can be evaluated using software for scientific computations. The pair-correlation function is defined as $g(\boldsymbol{s}_1, \boldsymbol{s}_2) = \frac{\rho_2(\boldsymbol{s}_1, \boldsymbol{s}_2)}{\rho_1(\boldsymbol{s}_1)\rho_1(\boldsymbol{s}_2)}$, and can hence be expressed using first and second product densities as given in Proposition 2.1.

Finally, the inhomogeneous empty space function [6, 17], $F(r)$, for the general model is given in the following proposition.

Proposition 2.2. *For an LSCP with log-intensity* (3) *where X_k are zero-mean stationary random fields with covariance functions $\mathcal{C}_k$, the inhomogeneous empty space function is given by*

$$F(\boldsymbol{s}_0, r) = 1 - \mathbb{E}\left[\prod_{k=1}^{K} \exp\left(-\int_{\mathcal{D}_k \cap B(\boldsymbol{s}_0, r)} e^{\mu_k(\boldsymbol{s})} e^{\mathrm{X}_k(\boldsymbol{s})} d\,\boldsymbol{s}\right)\right],$$

where for a given realization of X_0, $\mathcal{D}_k$ is the region classified as k.

The proof is given in Appendix B.

2.4 Finite dimensional approximation

As for standard LGCP models, some finite dimensional approximation of the LSCP model is needed if it is to be used for inference. The discretization we use here is a classical lattice approximation. The observational domain is discretized into N subregions $\mathcal{D}_i$ of a regular lattice over the domain, and the point locations are replaced by counts Y_i of the number of observations within each subregion $\mathcal{D}_i$. This yields the discretized model $\mathrm{Y}_i \sim \mathsf{Pois}(\lambda_i)$, where $\lambda_i = \int_{\mathcal{D}_j} \lambda(\boldsymbol{s}) d\boldsymbol{s}$ and the information on the fine-scale behaviour of the point pattern behavior is lost. The stochastic integral in the definition of λ_i is not Gaussian and generally difficult to handle. Therefore, a common approximation is to use $\lambda_j \approx |\, \mathcal{D}_j \,|\, \lambda(\boldsymbol{s}_j)$, for some location (usually the center) $\boldsymbol{s}_j \in \mathcal{D}_j$ [39].

For any fixed lattice approximation there is a positive probability that the level set field takes values in several of the intervals $\{c_k, c_{k+1}\}$ in any fixed lattice cell. Since the spatial information about the level set field on a finer scale than the lattice discretization are lost we propose adding a "nugget" effect, ξ_j, for each lattice cell $\mathcal{D}_j$. The "nugget" effect will model the within-cell classification uncertainty. This gives the discretization $\lambda_j \approx |\, \mathcal{D}_j \,| \tilde{\lambda}(\boldsymbol{s}_j)$, where

$$\log \tilde{\lambda}(\boldsymbol{s}_j) = \sum_{k=1}^{K} \mathbb{I}\left(c_{k-1} < X_0(\boldsymbol{s}_j) + \mu_0(\boldsymbol{s}_j) + \xi_j < c_k\right) \left(\mathrm{X}_k(\boldsymbol{s}_j) + \mu_k(\boldsymbol{s}_j)\right),$$

and $\xi_j \sim \mathsf{N}(0, {\sigma_\xi}^2)$. The *nugget variance*, ${\sigma_\xi}^2$, controls the amount of mixing between the classes for a given realization of X_0. This classification mechanism is equivalent to the ordered probit model discussed in Dunlop et al. [21]. In practice, it is typically difficult to objectively discern an appropriate value for σ_ξ and hence we therefore let σ_ξ be a regular parameter to be estimated for a fixed discretization.

In Appendix A, we show consistency of this finite dimensional approximation of the likelihood for the LSCP model. More precisely, we show that the posterior distribution for the latent fields $\{X_k\}_k$ computed using the lattice approximation converges, in total variation distance, to the posterior distribution of the continuous process.

3 Inference

A popular approach for Bayesian inference is through Markov chain Monte Carlo (MCMC) methodology, for instance using the Metropolis adjusted Langevin algorithm (MALA) [46] which was suggested for LGCP models by Møller et al. [39]. Another approach is through integrated nested Laplace approximation (INLA) [31, 48, 52], which when applicable can have beneficial computational properties. Unfortunately the level set construction cannot be handled by INLA.

In this work we propose an MCMC algorithm for fitting the LSCP model and perform predictions. Specifically, we propose a method based on the preconditioned Crank-Nicholson Langevin (pCNL) algorithm [16]. The algorithm is developed for the case when the target distribution (the posterior) has a Gaussian prior and a non-Gaussian likelihood, as is the case for the LSCP model. An important property of the pCNL algorithm is that it is discretization invariant [8]. This implies that as the finite dimensional approximation is made finer the mixing of the Markov Chain does not deteriorate. This is important since we utilize a rather fine discretization and thus one should expect poor mixing for an algorithm without this property, like for instance MALA.

We now present in more detail how we implement the MCMC algorithm for the LSCP model. Denote the set of parameters associated with class k as θ_k. For the level set field, X_0, we also include the nugget variance, σ_ξ, and the thresholds, $\{\mathrm{c}_k\}_k$ in the set θ_0. By introducing an auxiliary field Z defined such that $\mathbb{P}\left(\mathrm{Z}(\boldsymbol{s}_j) = k\right) = \Phi\left(\frac{\mathrm{c}_k - \mathrm{X}_0(\boldsymbol{s}_j)}{\sigma_\xi}\right) - \Phi\left(\frac{\mathrm{c}_{k-1} - \mathrm{X}_0(\boldsymbol{s}_j)}{\sigma_\xi}\right)$, we have

$$\log \tilde{\lambda}(\boldsymbol{s}) \stackrel{d}{=} \sum_{k=1}^{K} \mathbb{I}\left(\mathrm{Z}(\boldsymbol{s}) = k\right)\left(\mathrm{X}_k(\boldsymbol{s}) + \mu_k(\boldsymbol{s})\right).$$

This means that parameters and latent fields of different classes, $\{\mathrm{X}_k, \theta_k\}$, are conditionally independent given Z. We use this to construct a Metropolis-within-Gibbs algorithm [45] to sample from the joint posterior. In the ith iteration of the algorithm, the following three steps are performed

1. Sample from $\mathrm{Z}\,|\{\mathrm{X}_k, \theta_k\}_k, \mathrm{Y}$. The sampling can be performed exactly since $\mathrm{Z}(\boldsymbol{s}_i) \perp \mathrm{Z}(\boldsymbol{s}_j), \forall i \neq j$ given $\{\mathrm{X}_k, \theta_k\}_k, \mathrm{Y}$ and $\mathbb{P}\left(\mathrm{Z}(\boldsymbol{s}_i) = k\right)$ is known.

2. Sample from $\theta_k|\,\mathrm{Z}, \mathrm{X}_k$ using the MALA algorithm. Since parameters from different classes are conditionally independent, the sampling for each θ_k can be performed in parallell.

3. Sample from $\mathrm{X}_k\,|\,\mathrm{Z}, \theta_k, \mathrm{Y}$ using the pCNL algorithm. Also here, the updates for different k can be done in parallel since the fields, $\{X_k\}_{k=0}^K$, are conditionally independent.

The computational bottleneck of the algorithm is the third step, where the latent Gaussian fields are sampled. This is computationally expensive since the spatial discretization in two dimensions will correspond to a high dimensional distribution. However, since we are using a spatial lattice discretization with N subregions, efficient sampling from the prior Gaussian field is possible with a computational complexity of $\mathcal{O}(N \log N)$ using spectral methods [34] (that is, using Fast Fourier transforms). Working in the spectral domain also allows for efficient computation of all gradients and acceptance probabilities needed, making the spectral approach and the pCNL-algorithm in combination very favorable. Using the Fast Fourier transform relies on truncating the spectral basis expansion of the fields, In Appendix A we justify this additional approximation theoretically by showing that convergence of the lattice approximation still holds given certain bounds on the spectral densities.

4 Application

To further illustrate our approach we return to the tropical rainforest data example in Section 1 to compare the effect of considering LSCP models instead of simple Poisson regression or the LGCP model.

4.1 Data

The dataset consists of 2461 locations of trees of the species *Beilschmiedia pendula* in a 50 ha rectangular study plot (500 x 1000 meter) on the island of Barro Colorado in Panama, Figure 1a. The data were acquired from the first census of a major ongoing ecological study that started in the 1980s, designed to understand the mechanisms maintaining species richness, consisting of the observed positions of a large number of tree species ([28, 27, 13]). The study deliberately considers a spatially mapped rainforest community, arguing that population and community dynamics occur in a spatial context [26]. In addition to the spatial pattern formed by the tree locations, measurements of topographical variables and soil nutrients that potentially influence the spatial distribution of the trees are available [32, 51, 18], with the aim of linking spatial patterns to spatial environmental variations, reflected by observed topography and soil nutrients. Some of the point patterns derived from this material has been studied in, for instance, [38, 52, 31, 44] and the *Beilschmiedia pendula* data are available in the `spatstat` package [5] for the R project [1].

Elevation was measured and sampled on a 5x5 meter grid, and based on this an approximation of the slope at each of these grid points was calculated using a Sobel filter [54]. Soil samples were taken at 300 locations, for which the amount of 12 soil constituents (Al, B, Ca, Cu, Fe, K, Mg, Mn, N, Nmin, P, Zn) as well as the pH level were measured; these were interpolated to yield spatially continuous covariates. Since the covariates derived from the soil samples and elevation were not sampled at grid resolution they had to be interpolated to a common lattice. In this example, the model was discretized to 30×60 subregions over the observational window, giving a spatial resolution of 16.7×16.7 meters. The soil samples and elevation data was interpolated to the very same resolution. The number of observed points in each subregion is shown as a two dimensional histogram in Figure 1b. The spatial interpolation of the covariates to this lattice grid was performed using bi-cubic splines with the function *interp2* in Matlab (R2016a); Figure 3 shows the standardized covariates.

To avoid problems with multicollinearity among the covariates, covariates were discarded based on high variance inflation factors (VIF) [42]. The covariates were discarded iteratively by first computing the VIF for all covariates, removing the covariate corresponding to the highest VIF value if it exceeds 5 and then starting over on the new reduced set of covariates. The algorithm was stopped when none of the VIFs exceeded 5. By this procedure, the covariates B, Ca, K, and Zn were discarded, leaving 11 covariates for further analysis.

4.1.1 Models

It is obvious from Figure 1a that there is a large area in the middle of the domain where hardly any trees are growing. It is likely that some inhibitory factor prevents the trees from growing in that region. As mentioned earlier, we have anecdotal evidence that this area is covered by a swamp and that the tree species is known to be very unlikely to grow there. We test four different models to see how the confounding factor will affect inference., as sumarised in Table 1.

The first is a simple Poisson regression model on the covariates, i.e. an inhomogeneous Poisson process with linear fixed effects defining the log intensity as $\mathrm{B}(\boldsymbol{s})\boldsymbol{\beta}$. We will refer

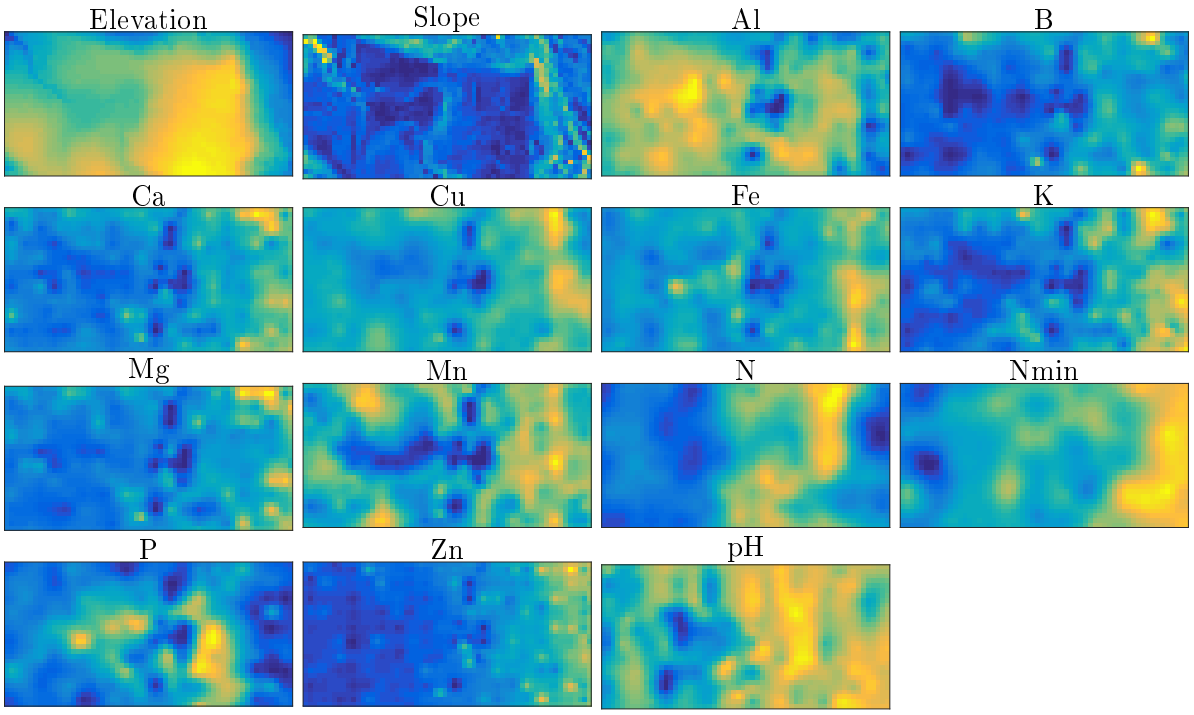


Figure 3: The standardized covariate values on the observational domain.

to this model as the *Fixed* model. The second model includes a Gaussian field to capture the variability not explained by the covariates. More precisely, we use a LGCP model with log-intensity $\log \lambda(\boldsymbol{s}) = \mathrm{X}(\boldsymbol{s}) + \mu(\boldsymbol{s})$. Here $\mathrm{X}(\boldsymbol{s})$ is a zero-mean Gaussian field with Matérn covariance with standard deviation σ and range r and $\mu(\boldsymbol{s}) = \mathrm{B}(\boldsymbol{s})\boldsymbol{\beta}$.

Looking at the data, we might expect the LGCP model to explain the variation in point intensity well, except for the complete lack of observations in the central region coupled with the discontinuity in the observed intensity at the border between the large empty area and the other parts of the plot. If the habitat dependence of the trees is significantly different in these two separated regions, a LSCP model with a separate class for each of the two regions might provide a better fit. Therefore, the third model is a two-class LSCP model where the first class is defined as in the LGCP model and the second class has a constant intensity fixed to a low value. That is, $\log \lambda(\boldsymbol{s}) = \mathrm{Z}_1(\boldsymbol{s})\,(\mathrm{X}_1(\boldsymbol{s}) + \mu_1(\boldsymbol{s})) + \mathrm{Z}_2(\boldsymbol{s})C_2$, where $\mathrm{Z}_k(\boldsymbol{s}) = \mathbb{I}\,(\mathrm{Z}(\boldsymbol{s}) = k)$. This choice of model is tailored to partition the spatial domain in to one partition with trees and one partition with a very small probability of trees growing. Since it is hard to distinguish between different values of very low intensity when studying a point pattern, the low intensity class might as well be modeled with a fixed intensity. We fix the intensity to $C_2 = 0.0268$, which is one tenth of the mean intensity among the grid cells with at most 1 tree. We refer to this as the *LGCPM* model. Finally, we consider a simplified version of this model where $\log \lambda(\boldsymbol{s}) = \mathrm{Z}_1(\boldsymbol{s})\,\mathrm{B}(\boldsymbol{s})\,\beta + \mathrm{Z}_2(\boldsymbol{s})C_2$. That is, no random effects in any of the two classes. This will be referred to as the *FixedM* model.

FIXED:	$\log \lambda(\boldsymbol{s}) = \mathrm{B}(\boldsymbol{s})\boldsymbol{\beta}$
LGCP:	$\log \lambda(\boldsymbol{s}) = \mathrm{X}(\boldsymbol{s}) + \mathrm{B}(\boldsymbol{s})\boldsymbol{\beta}$
LGCPM:	$\log \lambda(\boldsymbol{s}) = \mathrm{Z}_1(\boldsymbol{s})\,(\mathrm{X}(\boldsymbol{s}) + \mathrm{B}(\boldsymbol{s})\boldsymbol{\beta}) + \mathrm{Z}_2(\boldsymbol{s})C_2$
FIXEDM:	$\log \lambda(\boldsymbol{s}) = \mathrm{Z}_1(\boldsymbol{s})\,\mathrm{B}(\boldsymbol{s})\boldsymbol{\beta} + \mathrm{Z}_2(\boldsymbol{s})C_2$

Table 1: The four models and their corresponding log intensity functions.

The posterior distributions of the parameters and latent fields were estimated using the proposed MCMC method of Section 3. In order to reduce boundary effects, the lattice was extended by 350 m for the level set field and by 220 m for the latent Gaussian fields of the classes (implicitly assuming correlation ranges smaller than 350 m for classification and 220 m within classes). The smoothness parameters of the Gaussian fields were fixed at $\nu = 1$ and the following independent prior distributions for the model parameters (when applicable) were used: *i*) $\mathsf{N}(0, 10)$-priors for the fixed effects; *ii*) $\mathsf{N}(0, 4)$-priors for the threshold parameters; *iii*) an exponential prior with mean 2, Exp(2) for σ_k except that $\sigma_0 = 1$ is fixed due to the identifiability issues discussed in Section 2.1; *iv*) Exp(200)-priors truncated from below at the lattice distance and from above at the lattice extension range for r_k; this ensures that no wrap-around artifacts were introduced and that the correlation range was not smaller than the discretization distance; *v*) an Exp(0.1) prior truncated from above at 1 for the nugget standard deviation; this yields an expected *a priori* standard deviation of approximately 0.1 and ensures that the nugget variance does not dominate the spatial dependency in the level set field.

The standard deviations of the Gaussian fields were given exponential priors in order to penalize model complexity, following the PC prior framework of [55, 53]. This corresponds to penalization of the Gaussian fields, $\{\mathrm{X}_k\}_{k=1}^K$, when the covariates alone can explain the variation. The range parameters were given exponential priors using similar reasoning to the standard deviations where no spatial dependency corresponds to the base model. However, the distribution was truncated since no information exists on ranges below the lattice diameter due to the spatial discretization. The covariates were standardized to mean 0 and variance 1. Hence, the fixed effects prior yields a penalisation from the base model of no fixed effects. The nugget for the level set field was considered as a deviation from the base model (without a nugget) and hence penalised by an exponential distribution. The MCMC chains were run for $5 \cdot 10^6$ iterations on a standard contemporary laptop. The simulations took 2.3 hours for the Fixed model, 11 hours for the LGCP model, 32 hours for the LGCPM model, and 24 hours for the FixedM model.

4.1.2 Model validation

To validate the models we use a standard approach from point process literature [38, 30, 4] that compares a summary characteristic estimated from the observed point pattern with envelopes based on the summary characteristics estimated for point patterns simulated using the fitted models. As a summary characteristic we choose the pair correlation function [30] which is a second order functional summary characteristic characterizing the expected number of points at distances close to r from an observed point.

The pair correlation function with isotropic edge correction was estimated using the function *pcf* from the *spatstat* package [5]. An estimate from the observed data as well as estimates for each out of 5000 point pattern realizations was acquired for each model. Here we have first looked at point pattern realizations based on 5000 draws of parameter values $(\theta_0, \theta_1, \dots)$ from the posterior distribution given the observed data. That is, we compare the four models with parameter values that are reasonable given the observed trees data. Secondly we have also looked at 5000 point pattern realizations, each one sampled from an anisotropic Poisson process with an intensity function drawn from the posterior distribution conditioned on the observed data for each of the four models. That is, comparing point patterns generated from the posterior distributions of each of the four models. The draws from the posterior distributions were acquired using the MCMC simulations.

The pair correlation function for the observed point pattern as well as 5 randomly chosen realizations can be seen in Figure 4. The pointwise mean among all realizations and the 90% en-

velopes (both global and pointwise) have been plotted in the figure. The left column corresponds to realizations from the prior models using parameter values from the posterior distributions. The right column corresponds to realizations from the posterior point processes. For the Fixed model, the prior and posterior model do not differ since no random effects are present. Hence, only one figure is present in the corresponding column.

In general, the plots in the left hand (LH) column in Figure 4 reflect how well the simulated patterns resemble the overall pattern structure in the original data, and the plots on the right hand (RH) side how well the simulated patterns resemble the original pattern itself. The pair correlation function is defined for stationary patterns. In the current context, where we want to assess the models' ability to capture non-smooth non-stationarity in the pattern, the values of the summary characteristic might not be able to indicate issues with this type of structure. Hence, we also consider some randomly chosen examples of the corresponding simulated point patterns for each of the same scenarios as in Figure 4; these are shown in Figure 5. To assess the models' performance we consider these two figures together.

For the simplest model, the *Fixed* model, the estimated pcf for the observed point pattern is above the envelopes for all distances (Figure 4). This clearly indicates that the model that is based on covariates only does not sufficiently capture aggregation at any spatial scale and is clearly inappropriate for this data set. This is not surprising since it had been anticipated that the covariates are unable to reflect the spatial structure resulting from the presence/absence of the swamp. In addition, they might also not be sufficient to capture small scale clustering, resulting from dispersal limitation. The relevant simulated pattern in Figure 5 reflects this very clearly; the pattern does not show any of the small scale clustering that's evident in the observed pattern and it does not sufficiently explain the large empty area where the swamp is located. There is an area with lower point intensity, perhaps a result of some of the covariates being correlated with the presence of the swamp.

For the standard LGCP model, the estimated pair correlation functions in Figure 4 (LH and RH) for the observed point pattern remain inside the global envelopes. However, for small distances, the mean function for the patterns simulated from the fitted model is above the function for the observed pattern in the LH plot. This is reflected in the low values of the black line relative to the red line for up to about 40 meters radius r. This indicates that the model accounts for less clustering at small distances than has been observed in the pattern formed by the rainforest trees. This is likely a result of the model attempting to average the parameters for the Gaussian field across the whole plot; the Gaussian field in the model attempts to explain spatial structures that the covariates cannot explain in general as well as the the effect of the swamp. This is also evident in the relevant LH plot in Figure 5. The pattern shows some empty areas, perhaps smaller than the swamp, but this is hard to discern from a single pattern only. It also shows local clustering which is much stronger than in the original pattern. The associated RH plot looks rather similar to the observed pattern with perhaps too many points in the clustered areas.

For the LGCPM model, the mean line and the estimated line pair correlation functions are very similar, even though the red line is slightly above the black line for small distances in the LH plot of Figure 4. This might indicate that slightly more local clustering is reflected in the model than is exhibited by the actual pattern. The general spatial structure in the LH simulated pattern appears to be much more similar to the general structure in the original pattern than that simulated from the LGCP model, neither overly exaggerating the clustering nor the "emptiness". The RH equivalent again strongly resembles the original pattern, perhaps with slightly too many points in the areas with local clustering as indicated by the pair correlation function.

The FixedM model underestimates local clustering, most likely behavior resulting from dispersal limitation, which is not accounted for by the covariates. This becomes most apparent

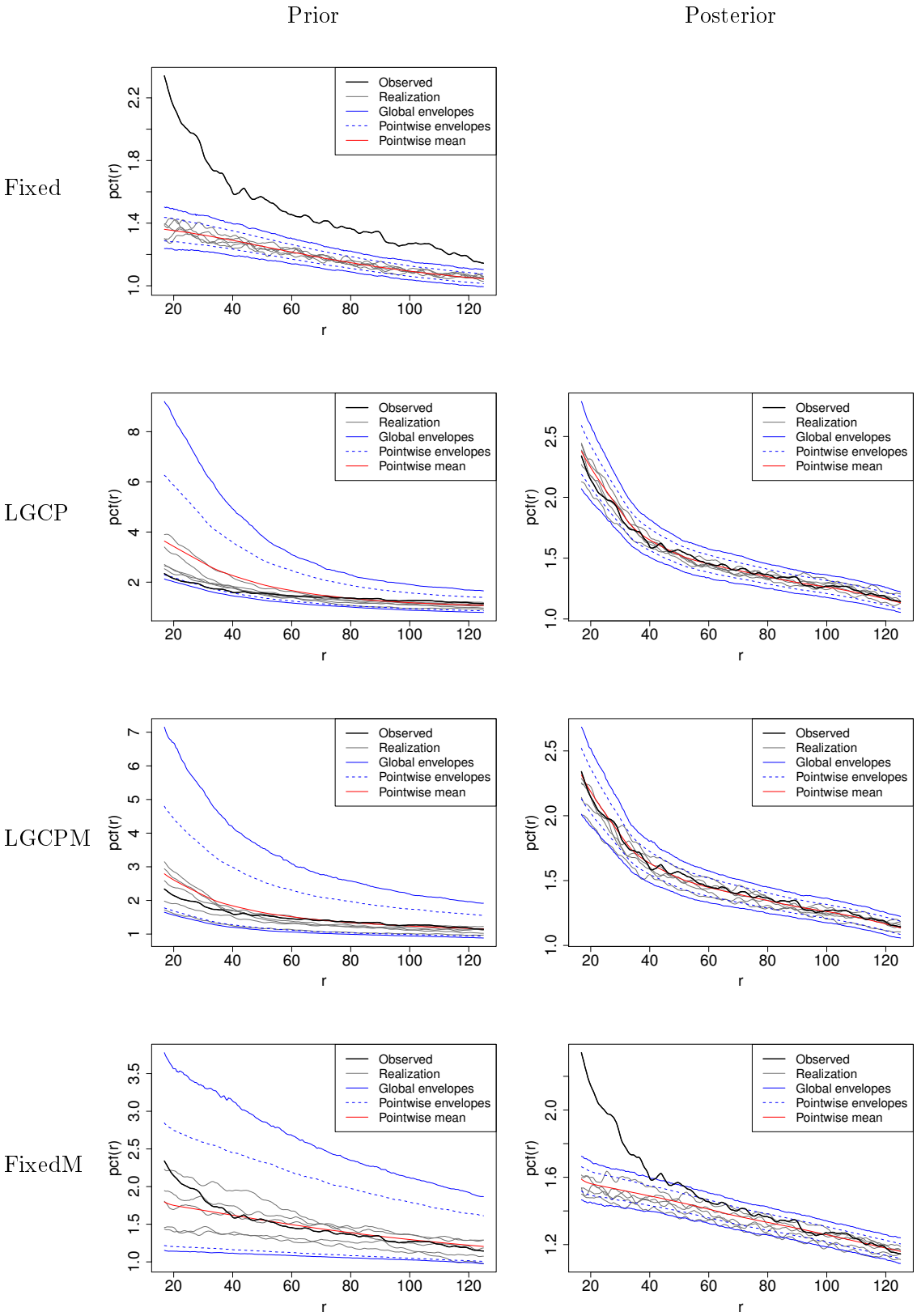


Figure 4: Plots of the pair correlation function. Left column uses envelopes and realizations from point patterns generated using the prior model with parameters drawn from the posterior distribution. Right column uses envelopes and realizations drawn from the posterior point processes. Each panel show the value for the observed point pattern (black), 5 randomly chosen realizations (gray), the pointwise mean (red), 90% pointwise (blue dotted) and global (blue) envelopes.

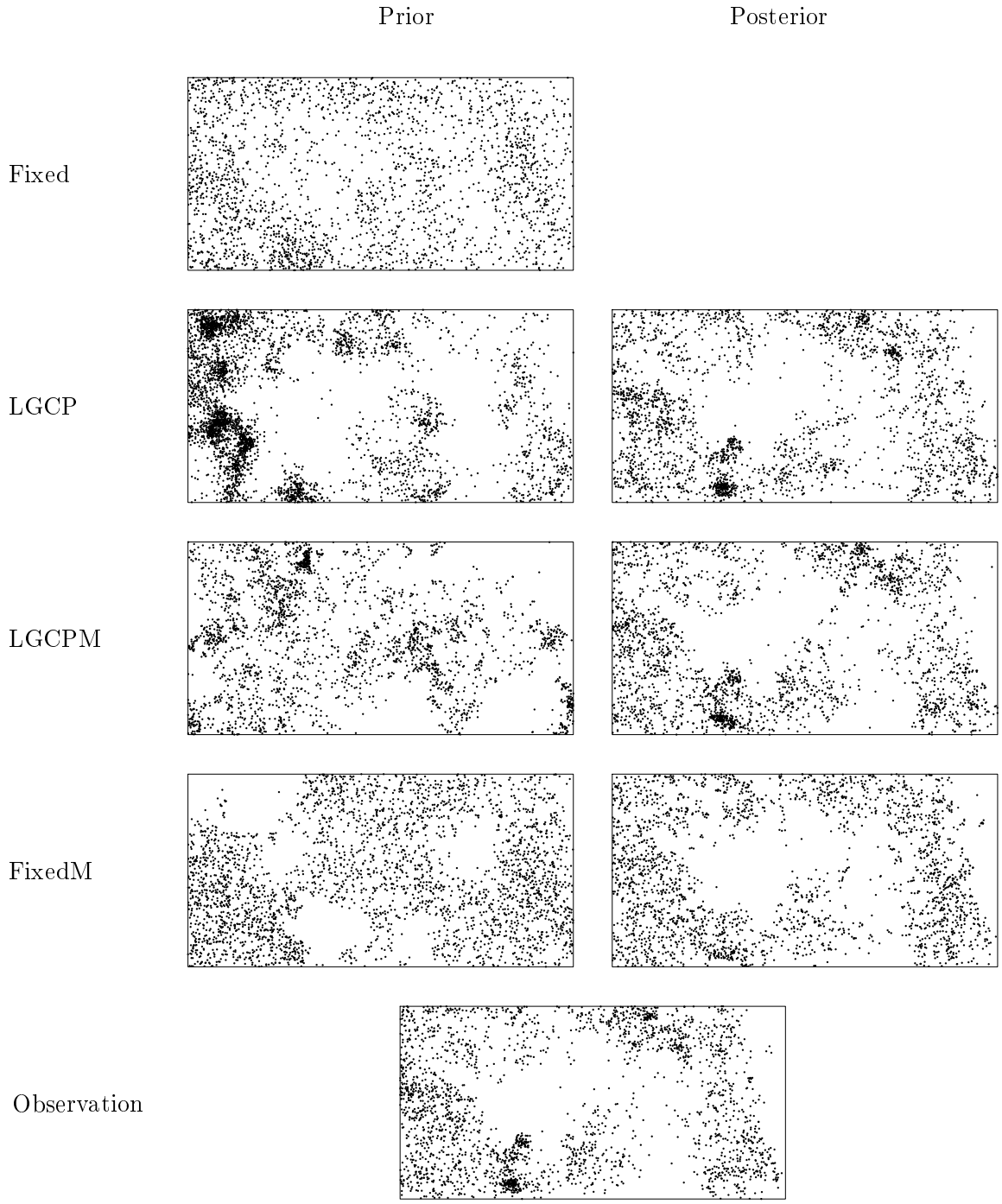


Figure 5: Example of point patterns generated from all four models. Left column corresponds to realization from prior model but with parameter values drawn from the posterior distribution. Right column are point patterns generated from the posterior point processes.

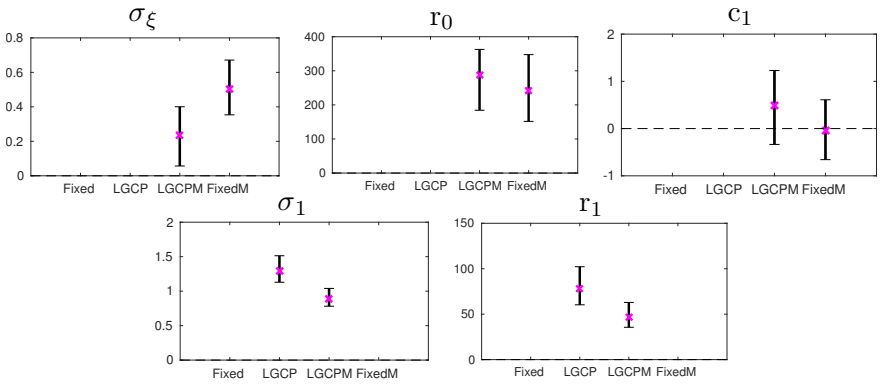


Figure 6: The mean (cross) and 95% credible intervals (lines) for the field parameters.

in the RH plot of Figure 4, where the function estimated for the observed pattern is above the envelopes up to a distance of about 40 meters, but is also evident in the associated LH plot. The simulated patterns in Figure 5 tell a similar story; the swamp is reasonably well accounted for but there is a clear lack of local overdispersion in the patterns.

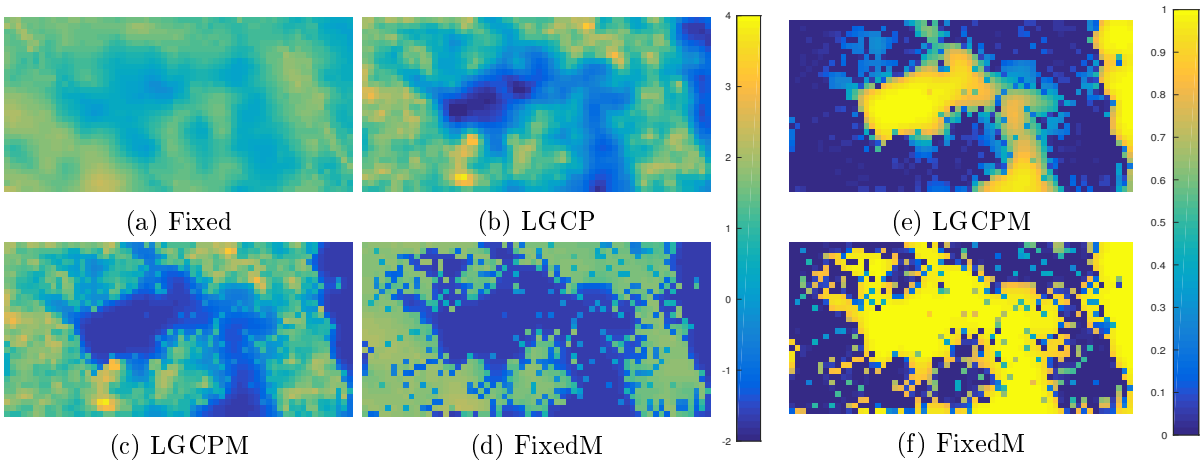
4.1.3 Analysis of covariates and spatial structure

The models discussed here, relating a spatial pattern to spatially continuous covariates may be of interest for a number of reasons. Commonly, one seeks to understand habitat preferences of a particular species as reflected in the relationship between the point pattern and the covariates. In addition, it might be of interest to understand the nature of the spatial structure that remains unexplained by the covariates.

To investigate the spatial structure, we first look at the mean value and 95% credible intervals for the random field parameters of the models. These are presented in Figure 6. Observing the difference between r_1, and σ_1 values of the LGCP and LGCPM models show how the empty region will affect the estimation of the spatial dependency structure. Here, the LGCPM model shows a significantly lower variance and clearly lower correlation range. This is natural since the Gaussian field for the LGCPM model does not need to explain both the effect of natural spatial dependency between growth of trees as well as the unknown inhibitory effect that causes trees to not grow at all in certain regions of the forest. And finally we note that σ_ϵ has a large effect (signal to noise ratio equals $\frac{1}{\sigma_\epsilon}$) indicating that the Matérn field for X_0 cannot explain the classification on its own. This is clearer for the FixedM model, where classification jumps more sporadically between adjacent grid cells due to the over-simplified structure of the classes.

In Figure 7 the mean posterior log intensities, $\{\log \lambda_i\}_{i=1}^N$ are presented as kriging predictions for each of the four models. The figure also shows the posterior probabilities $\mathbb{P}\left(\mathrm{Z}(\boldsymbol{s}) = 2|\mathrm{Y}\right)$, giving an indication of the region with very few trees. The posterior log intensity surface of the LGCPM shows sharp boundaries contrary to the smoothly varying in the LGCP. The classification in the FixedM model is more noisy than that of the LGCPM model and a larger proportion of the observation window is classified as being the *empty region*. Once again, this is connected to the larger value of σ_ξ and caused by the FixedM having to explain the intensity with a much simpler model.

The relationship between the tree intensity and the covariates is also of interest, and in practice is often the focus of a study. Recall that 12 covariates of the original 16 covariates are considered here; 11 true covariates and one intercept term. Figure 8 shows the mean and 95% credible intervals for each of these covariates for all models. The first question is which of the covariates have a significant impact on the spatial distribution of the trees and hence reflect a



(a) Fixed (b) LGCP (e) LGCPM

(c) LGCPM (d) FixedM (f) FixedM

Figure 7: Mean posterior log intensity surface, λ (left) and mean classification for the models where applicable (right)

Model	Covariates
Fixed	Int, Elev, Slope, Al, Mn, NMin, P, pH, Cu, N
LGCP	Int
LGCPM	Int
FixedM	Int, NMin, Elev, Mn

Table 2: Significant covariates on a 5% level for the covariates using Holm-Bonferroni correction to correct for multiple hypothesis tests.

habitat preference of the species. To answer this we asses which of the regression coefficients in $\boldsymbol{\beta}$ that are significantly different from zero. Empirical p-values are computed from the sampled posterior distributions. Since there are 12 fixed effects, the p-values were adjusted using Holm-Bonferroni correction [25] to account for multiple testing. Table 2 shows the covariates that were considered significant, at a significance level of 5%, for each of the four models.

The FixedM model identifies a smaller number of significant covariates than the Fixed model. This is not surprising since the covariates do not need to explain the lack of trees in the empty domain anymore. It should be noted that we found that the Fixed and FixedM models did not fit the data, see Figure 4. Hence, finding significant covariates could be due to model misspecification rather than an actual relationship with the observed point pattern. For the LGCP and LGCPM models, only the intercepts are significant.

5 Discussion

We have considered the problem of Bayesian level set inversion for point process data. The proposed model can be seen as a generalization of the LGCP model where the latent Gaussian field is extended to a level set mixture of Gaussian fields. We derived basic model properties and in Appendix A showed consistency of the posterior probability measure of finite-dimensional approximations to the continuous model. A computationally efficient MCMC method for Bayesian inference, based on the preconditioned Crank-Nicholson Langevin algorithm, was presented. A topic of further research could be to investigate other, potentially even quicker, estimation methods such as based on INLA in combination with variational Bayes.

We modelled a point pattern formed by the locations of the trees from a species in a tropical

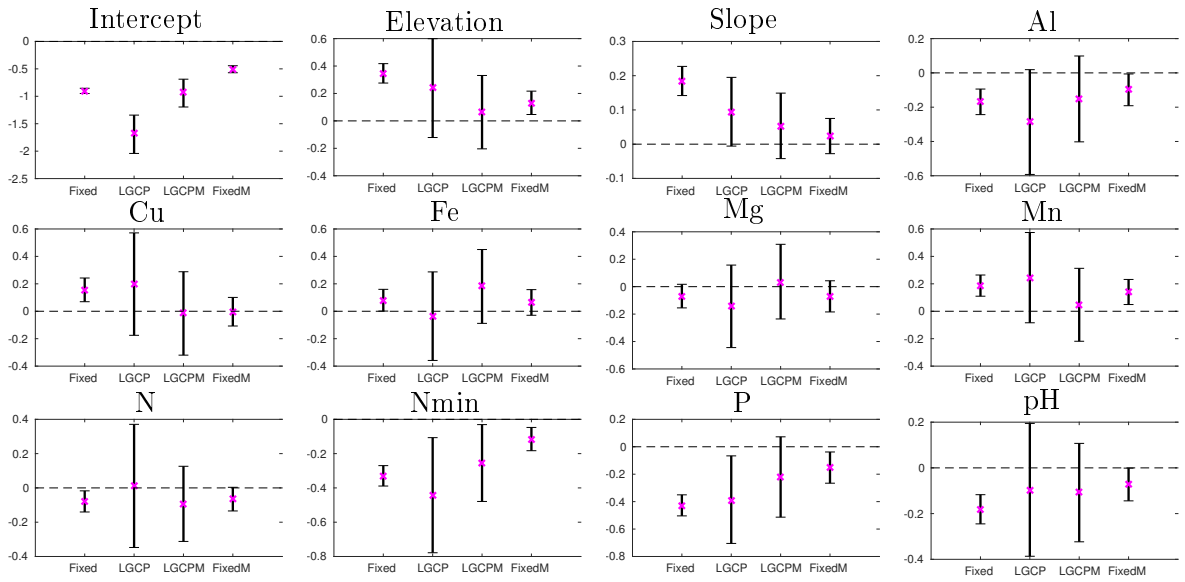


Figure 8: The mean (cross) and 95% credible intervals (lines) for the posterior marginal distribution of fixed effect for the four different models.

rainforest. The example was relevant in the current context since the point pattern shows clear signs of being affected by some confounding factor (mainly reflecting the presence or absence of a swamp). Soil and topography covariates are at most slightly correlated with the swamp and hence cannot explain the nearly complete absence of points in that area. A standard LGCP model accounts for remaining spatial structures through a single smooth Gaussian field. The LSCP model we propose here can have a level set mixture of two Gaussian fields, one component explaining local spatial structures (such as those resulting from dispersal limitation) and another constant component explaining the absence or presence of the "emptiness".

We fitted 4 different models to the data, which all could be viewed as special cases of the LSCP models, a simple model with only covariates (Fixed), a standard LGCP process model, a LSCP model with one constant class and one class modeled using a Gaussian field (LGCPM) and a simplified version of this which only had fixed effects in the first class (FixedM). We compared the performance of the four models using simulations from the fitted models to the observed data, both based on summary characteristics from repeated simulation and by visually comparing the resulting spatial structures in the simulated patterns to those in the observed pattern. The differences in model performance were most evident in the visual comparison of the generated patterns. The LGCPM clearly reproduced the original pattern structure better than the other models, with the Fixed model neither showing local clustering nor a meaningful empty area, the LGCP model exhibiting exaggerated local clustering and the FixedM model not accounting for local clustering. The pair correlation function reflected these features as well, but not in as an obvious way. The analysis of the tropical rainforest showed that inference on both the Gaussian field parameters and covariates were affected by allowing for a second class in the model. It suggests that the inference drawn based on the LGCP model were biased by the confounding factor and that no covariates can be considered statistically significant.

Patterns simulated from the model might also be useful for aiding the interpretation and understanding of underlying ecological mechanisms. For instance, simulating patterns based on the parameters estimated from the LGCPM but not involving the component relating to the swamp in the simulation – as it is clear that it leads to low point intensities – would provide an insight into what spatial structures looked like if the swamp was not present. This is interesting,

as the "swamp effect" trivially leads to spatial structures that are uninteresting but very obvious on visual inspection.

Future analysis could consider using fixed effects also with the level set field, X_0, in order to investigate which covariates explain the classification, a feature of the proposed model that we have not yet investigated. Further, multi-type point patterns may also be analyzed with the proposed model class for instance for the joint analysis of several species of plants. This could be performed by introducing multivariate Gaussian random fields for the classes, i.e. for $\{\mathrm{X}_k\}_{k=1}^K$. Another possibility is letting several species share the same level set field, X_0, or classifications field, Z, but use independent class fields, $\{\mathrm{X}_k\}_{k=1}^K$. In this way, information about X_0 could be pooled from several point patterns jointly.

Furthermore, the approach may also be relevant in larger scale studies, for example in the context of species distribution modelling, where data collection effort varies in space and whole spatial areas have not been surveyed or only very few surveys have taken place, despite providing potentially suitable habitat for a species. This would create large areas relatively empty of points for which spatial covariates are available. In particular, if suitable covariates were available that might be linked to survey effort, such as population density or accessibility, a suitable level set mixture model could account for this.

A problem with the spectral approach used in this work is that the spatial discretization has to be on a lattice. In applications where such restrictions are problematic, the sampling of the Gaussian fields could be performed with a different method. Generally this requires (KN^3) operations. A possible approach to remedy this would be to acquire a Gaussian Markov random field approximation of the problem. This idea has been studied by [35, 47, 52] revealing computationally attractive properties on arbitrary domains. An adaptation of the method by Simpson et al. [52] to the LSCP model would reduce the computational cost to $O(KN^{3/2})$ while still allowing for arbitrary spatial discretizations.

Another issues that needs further investigation relates to the prior choice for the parameters of the Gaussian random field as these can substantially influence the smoothness and, as a result, the significance of the spatial covariates. A spatial field that is too "wiggly" can easily lead to overfitting, rendering any covariates insignificant, while a field that is too smooth defies its very purpose. Some work has been done on investigating this in the context of pc priors for log Gaussian Cox processes [55], but there is room for further investigation.

6 Acknowledgements

The authors gratefully acknowledge the financial support from the Knut and Alice Wallenberg Foundation, the Swedish Research Council Grant 2016-04187, and the ÅForsk foundation. We would like to thank the people at the Center of tropical forest research, Smithsonian Tropical Research Institute for the extensive forest census plot and for making the data publicly available. The BCI forest dynamics research project was founded by S.P. Hubbell and R.B. Foster and is now managed by R. Condit, S. Lao, and R. Perez under the Center for Tropical Forest Science and the Smithsonian Tropical Research in Panama. Numerous organizations have provided funding, principally the U.S. National Science Foundation, and hundreds of field workers have contributed.

Also thanks to the Barro Colorado soil survey (Jim Dalling, Robert John, Kyle Harms, Robert Stallard and Joe Yavitt and field assistants Paolo Segre and Juan Di Trani) for making the soil sample data publicly available and for answering questions and handing out the original soil sample locations on request. The Barro Colorado soil survey was funded by NSF DEB021104,021115, 0212284,0212818 and OISE 0314581 as well as the STRI Soils Initiative and CTFS.

References

[1] The R Project for Statistical Computing. URL `https://www.r-project.org/`.

[2] R.J. Adler and J.E. Taylor. *Random Fields and Geometry*. Springer, 2007. ISBN 978-0-387-48112-8.

[3] J-M. Azaïs and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields*. Wiley, 1 edition, 2009. ISBN 04704093393.

[4] A. Baddeley, E. Rubak, and R. Turner. *Spatial point patterns: methodology and applications with R*. CRC Press, 2015.

[5] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015. URL `http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/`.

[6] A.J. Baddeley. Non- and semi-parametric estimation of interaction in inhomogeneuous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000.

[7] S. Barman and D. Bolin. A three-dimensional statistical model for imaged microstructures of porous polymer films. *Journal of Microscopy*, 269(3):247–258, 2018.

[8] Alexandros Beskos, Gareth Roberts, Andrew M Stuart, and Jochen Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8(3):319–350, 2008.

[9] M. Burger. A level set method for inverse problems. *Inverse problems*, 17(5):1327–1355, 2001.

[10] D. F. R. P. Burslem, N. C. Garwood, and S. C. Thomas. Tropical forest diversity – the plot thickens. *Science*, 291:606–607, 2001.

[11] E.T. Chung. Electrical impedance tomography using level set representation and total variational regularization. *Journal of computational physics*, 205:357–372, 2005.

[12] R. Condit. *Tropical Forest Census Plots*. Springer-Verlag and R. G. Landes Company, Berlin, Germany, and Georgetown, Texas., 1998.

[13] R. Condit. *Tropical Forest Census Plots: Methods and Results from Barro Colorado Island, Panama and a Comparison with Other Plots*. Springer Berlin Heidelberg, 1998. ISBN 9783662036648.

[14] R. Condit, P. S. Ashton, P. Baker, S. Bunyavejchewin, S. Gunatilleke, N. Gunatilleke, S.P. Hubbell, R.B. Foster, A. Itoh, J.V. LaFrankie, H.S. Lee, E. Losos, N. Manokaran, R. Sukumar, and T. Yamakura. Spatial patterns in the distribution of tropical tree species. *Science*, 288:1414–1418, 2000.

[15] S.L. Cotter, M. Dashti, and A.M. Stuart. Approximation of Bayesian inverse problems for PDEs. *SIAM journal on numerical analysis*, 48(1):322–345, 2010.

[16] S.L. Cotter, G.O. Roberts, A.M. Stuart, and D. White. MCMC Methods for Functions: Modifying Old Algorithms to Make Them Faster. *Statistical Science*, 28(3):424–446, 2013.

[17] D.J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes: Volume II: General Theory and Structure*, volume 2. Springer, 2003. ISBN 0-387-95541-0.

[18] J. Dalling, R. John, K. Harms, R. Stallard, and J. Yavitt. Soil Maps of Barro Colorado Island 50 ha Plot. URL "http://ctfs.si.edu/webatlas/datasets/bci/soilmaps/BCIsoil.html".

[19] O. Desjardins and H. Pitsch. A spectrally refined interface approach for simulating multiphase flows. *Journal of computational physics*, 228(5):1658–1677, 2009.

[20] P.J. Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. Third edition edition, 2014.

[21] M.M. Dunlop, M.A. Iglesias, and A.M. Stuart. Hierarchical Bayesian level set inversion. *Statistics and Computing*, pages 1–30, 2016.

[22] M. Fuentes. A new class of nonstationary spatial models. Unpublished manuscript, available at "http://www.stat.unc.edu/postscript/rs/nonstat.pdf", 2001.

[23] J. Hendrix, T. Dekens, W. Schrimpf, and D.C. Lamb. Arbitrary-Region Raster Image Correlation Spectroscopy. *Biophysical Journal*, 111(8):1785–1796, 2016.

[24] M.D. Higgs and J.A. Hoeting. A clipped latent variable model for spatially correlated ordered categorical data. *Computational statistics and data analysis*, 54:1999–2011, 2010.

[25] S. Holm. A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian journal of statistics*, 6(2):65–70, 1979.

[26] S. P. Hubbell. *The Unified Neutral Theory of Biodiversity and Biogeography*. Monographs in Population Biology 32, Princeton University Press, 2001.

[27] S. P. Hubbell, R. B. Foster, S. T. O'Brien, K. E. Harms, R. Condit, B. Wechsler, S. J. Wright, and S. Loo de Lao. Light-Gap Disturbances, Recruitment Limitation, and Tree Diversity in a Neotropical Forest. *Science*, 283(5401):554–557, 1999.

[28] S. P. Hubbell, R. Condit, and R. B. Foster. Barro Colorado Forest Census Plot Data, 2005. URL http://ctfs.si/edu/datasets/bci.

[29] M.A. Iglesias, Y. Lu, and A.M. Stuart. A Bayesian level set method for geometric inverse problems. *Interfaces and free boundaries*, 18(2):181–217, 2016.

[30] J. B. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*, volume 70. John Wiley & Sons, 2008.

[31] J.B. Illian, S.H. Sørbye, and H. Rue. A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation. *The Annals of Applied Statistics*, 6 (4):1499–1530, 2012.

[32] R. C. John, J. W. Dalling, K. E. Harms, J. B. Yavitt, R. F. Stallard, M. Mirabello, S. P. Hubbell, R. Valencia, H. Navarrete, M. Vallejo, and R. B. Foster. Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences USA*, 104:864–869, 2007.

[33] D. Lambert. Zero-Inflated Poisson Regression, With an Application to Defects in Manufacturing. *Technometrics*, 34(1):1–14, 1992.

[34] A. Lang and J. Potthoff. Fast simulation of Gaussian random fields. *Monte Carlo Methods and Applications*, 17(3):195–214, 2011.

[35] F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society*, 73(4):423–498, 2011.

[36] R.J. Lorentzen, G. Naevdal, and A. Shafieirad. Estimating Facies Fields by Use of the Ensemble Kalman Filter and Distance Functions - Applied to Shallow-Marine Environments. *SPE Journal*, 3(1):146–158, 2012.

[37] B. Matérn. *Spatial Variations*, volume 36. Springer-Verlag, 1986. ISBN 9780387963655.

[38] J Møller and R.P. Waagepetersen. Modern Statistics for Spatial Point Processes. *Scandinavian Journal of Statistics*, 34(4):643–684, 2007.

[39] J Møller, A.R. Syversveen, and R.P. Waagepetersen. Log Gaussian Cox Processes. *Scandinavian journal of statistics*, 25(3):451–482, 1998.

[40] V.V. Mourzenko. Percolation in two-scale porous media. *The European physical journal B*, 19(1):75–85, 2001.

[41] M. Myllymäki and A. Penttinen. Bayesian inference for Gaussian excursion set generated Cox processes with set-marking. *Statistics and computing*, 20(3):305–315, 2010.

[42] J. Neter, W. Wasserman, and M.H. Kutner. *Applied Linear Regression Models*. Irwin, second edition edition, 1989. ISBN 0-256-07068-7.

[43] D.B. Owen. A table of normal integrals. *Communication in Statistics - Simulations and Computation*, 9(4):389–419, 1980.

[44] T. Rajala and J. Illian. A family of spatial biodiversity measures based on graphs. *Environmental and ecological statistics*, 19(4):545–572, 2012.

[45] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer, 2 edition, 2004. ISBN 9781475741452.

[46] G.O. Roberts and R.L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996.

[47] H. Rue and L. Held. *Gaussian Markov random fields*, volume 104. Chapman and Hall, 2005. ISBN 0203492021.

[48] H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: series B*, 71(2):319–392, 2009.

[49] F. Santosa. A level-set approach for inverse problems involving obstacles. *ESAIM: Control, Optimisation and caculus of variations*, 1:17–33, 1996.

[50] B. Scheuermann and B. Rosenhahn. Analysis of Numerical Methods for Level Set Based Image Segmentation. *Lecture notes in computer science*, 5876(2):196–207, 2009.

[51] L. A. Schreeg, W. J. Kress, D. L. Erickson, and N. G. Swenson. Phylogenetic analysis of local-scale tree soil associations in a lowland moist tropical forest. *PLoS ONE*, 5:1–10, 2010.

[52] D. Simpson, J.B. Illian, F. Lindgren, S.H. Sørbye, and H. Rue. Going off grid: computational efficient inference for log-Gaussian Cox processes. *Biometrika*, 103(1):49–70, 2016.

[53] D. Simpson, H. Rue, A. Riebler, T.G. Martins, and Sørbye S.H. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical science*, 32(1):1–28, 2017.

[54] M Sonka, V Hlavac, and R Boyle. *Image Processing, Analysis, and Machine Vision*, chapter Image pre-processing. Thomson, 2008.

[55] S. H. Sørbye, J. B. Illian, D. P. Simpson, D. Burslem, and H. Rue. Careful prior specification avoids incautious inference for log-gaussian cox point processes. *arXiv preprint arXiv:1709.06781*, 2017.

[56] A.M. Stuart. Inverse problems: A Bayesian perspective. *Acta numerica*, 19:451–559, 2010.

A Theoretical results

In this section, we will theoretically justify the two approximations of the LSCP process that are needed for inference. The first is the finite dimensional approximation from Section 2.4 and the second is the truncation needed for the fast Fourier transform in Section 3.

For $k = \{0, ..., K\}$, let X_k be a Gaussian random field on the spatial domain $\mathcal{D} = [0,1]^d \subset \mathbb{R}^d$, defined on a complete probability space. We will show the results using methods similar to those in [15, 29, 52] and for this it is convenient to represent the fields as Gaussian measures $\mu_0^{(k)}$. To simplify the presentation, we will assume a specific covariance operator, $\mathcal{C}$, related to the Matérn covariance function. However, the results can be extended to more general densely-defined, self-adjoint, positive definite operators and to more general bounded domains.

Let $\mu_0^{(k)} = \mathcal{N}(0, \mathcal{C})$, where $\mathcal{C} = \tau^2 A^{-\alpha}$ with $A = \kappa^2 - \Delta$. Here τ, κ^2 and α are positive parameters and $A : \mathscr{D}(A) \subset L_2(\mathcal{D}) \to L_2(\mathcal{D})$, here $\mathscr{D}(A)$ denotes the domain of A. Further we impose periodic boundary conditions. Denote the eigenvalues of A as $\{\lambda_j\}_{j\in\mathbb{N}}$, which are arranged in a nondecreasing order, and the corresponding eigenfunctions as $\{e_j\}_{j\in\mathbb{N}}$, which form a complete orthonormal basis for $L_2(\mathcal{D})$. The fractional power operator $A^\alpha : \mathscr{D}(A^\alpha) \to L_2(\mathcal{D})$ is defined by

$$A^\alpha u = \sum_{j\in\mathbb{N}} \lambda_j^\alpha \langle u, e_j \rangle e_j.$$

For any α, the subspace $\mathcal{H}^\alpha := \mathscr{D}(A^{\alpha/2})$ is a Hilbert space

$$\mathcal{H}^\alpha = \{u : \sum_{j\in\mathbb{N}} \lambda_j^\alpha |\langle u, e_j \rangle|^2 < \infty\},$$

with respect to the inner product $\langle \phi, \psi \rangle_\alpha = \left\langle A^{\alpha/2}\phi, A^{\alpha/2}\psi \right\rangle$ and corresponding norm $\|\phi\|_\alpha = \sum_{j\in\mathbb{N}} \lambda_j^\alpha \langle \phi, e_j \rangle^2$.

With this choice of covariance operator, we have that if $u \sim \mu_0^k$, then $u \in \mathcal{H}^s$ for any $s < \alpha - d/2$ μ_0^k-almost surely [21, Theorem 1]. Furthermore, u is almost surely p-times differentiable if $\alpha - d/2 > p$. We will need this differentiability and we formulate it as an assumption.

Assumption A.1. *The classification field* X_0 *is almost surely a Morse function with strictly positive variance at all locations in the domain, and for* $k > 0$ *the Gaussian fields* X_k *are almost surely differentiable.*

The differentiability assumption is satisfied by assuming $\alpha > 2$. The Morse function requirement is slightly stronger than C^2, but is implied by $\alpha > 4$ [2]. Furthermore, we can use

a theorem equivalent to the Sobolev embedding theorem for our $\mathcal{H}^s$ space [56, Theorem 2.10]. That is, $\|X_k\|_{L^\infty} \leq C\|X_k\|_s$ if $X_k \in \mathcal{H}^s$ and $s > d/2$. For our case with periodic boundary conditions the space $\mathcal{H}^s$ is even equivalent to the Sobolev space H^s.

We thus have that X_k is represented as a Gaussian measure, $\mu_0^{(k)}$, on $\mathcal{H}^\alpha$ and we can choose an appropriate σ-algebra such as the probability space $(\mathcal{H}^\alpha, \Sigma_k, \mu_0^{(k)})$ becomes complete (see [29]). Likewise $\mathrm{X} = \{\mathrm{X}\}_{k=0}^K$ can be represented by a product measure μ_0 on the complete measure space $\mathcal{X} = (\Omega, \Sigma, \mu_0)$, where Ω is the product space of each $\mathcal{H}^\alpha$ and Σ is the corresponding product σ-algebra.

Since the LSCP model defines the point process as a non-homogeneous Poisson process conditioned on X, the likelihood potentials for the continuous and finite dimensional models, defined in Section 2.4, are

$$\Phi(\mathrm{X};\mathrm{Y}) = \int_{\mathcal{D}} \lambda(s;\mathrm{X})ds - \sum_{\boldsymbol{s}_j \in \mathrm{Y}} \log \lambda(\boldsymbol{s}_j;\mathrm{X}), \tag{4}$$

$$\Phi^N(\mathrm{X};\mathrm{Y}) = \sum_{i \in N} \left(|\, \mathcal{D}_i \,|\, \lambda(\tilde{\boldsymbol{s}}_i;\mathrm{X}) - \mathrm{Y}_i \log \lambda(\tilde{\boldsymbol{s}}_i;\mathrm{X}) \right). \tag{5}$$

Here, N is the number of discretized regions in the lattice approximation and Y_i denotes the number of observations in $\mathcal{D}_i$. Further, $\tilde{\boldsymbol{s}}_i$ is the midpoint of each $\mathcal{D}_i$, and $\boldsymbol{s}_j$ is the location of the jth point in the point pattern Y. Based on these likelihoods, we can now define the corresponding posterior measures as follows.

Proposition A.2. *If Assumption A.1 holds, we can define posterior measures using Radon-Nikodym derivative with respect to μ_0:*

$$\begin{aligned} \frac{d\mu}{d\mu_0}(\mathrm{X}) &= \frac{1}{C_\mu(\mathrm{Y})} \exp\left(-\Phi(\mathrm{X};\mathrm{Y})\right), \\ \frac{d\mu^N}{d\mu_0}(\mathrm{X}) &= \frac{1}{C_{\mu^N}(\mathrm{Y})} \exp\left(-\Phi^N(\mathrm{X};\mathrm{Y})\right), \end{aligned} \tag{6}$$

where $C_\mu(\mathrm{Y})$ and $C_{\mu^N}(\mathrm{Y})$ are normalizing constants.

The proof is given in Appendix B. Since only the discretized model can be used for inference, it is important to know that the approximation μ^N converges to the true posterior, μ, as the discretization becomes finer. The following theorem shows that this indeed is the case with respect to the total variation distance, $d_{\mathrm{TV}}(\mu, \mu^N) = 2\sup_{E \in \mathcal{F}_\mathcal{X}} |\mu(E) - \mu^N(E)|$.

Theorem A.3. *Let Assumption A.1 hold and let μ^N and μ be the posterior measures defined in (6). Then $d_{TV}(\mu, \mu^N) \to 0$ as $N \to \infty$.*

The proof is given in Appendix B. Also the latent fields, X, need to be approximated by finite dimensional representations for inference. We will do this by truncating the basis expansion of the field to p terms:

$$\mathrm{X} \approx \tilde{X} = \sum_{j=1}^{p} \xi_j \lambda_j^\alpha e_j,$$

where ξ_j are independent standard normal variables. We will refer to the model using a discretization of the observational domain and finite dimensional approximations of X as the *fully discretized model*. The advantage with using this truncation is that we can use the fast Fourier transform for simulating the field. To show that we still have convergence under this approximations, note that the finite dimensional approximation of X can be viewed as an orthogonal

projection of X on to the space spanned by the eigenfunctions $\{e_j\}_{j\leq p}$ as is done in Cotter et al. [15]. We define the projection operator P^p such that $\tilde{\mathrm{X}}(\boldsymbol{s}) = P^p\,\mathrm{X}(\boldsymbol{s})$. It is now possible to define a posterior probability measure for $\tilde{\mu}^N$ by it's Radon-Nikodym derivative as

$$\frac{d\tilde{\mu}^N}{d\mu_0}(\mathrm{X}) = \frac{1}{C_{\tilde{\mu}^N}(\mathrm{Y})}\exp\left(-\Phi^N(P^p\,\mathrm{X};\mathrm{Y})\right). \tag{7}$$

An important consequence of this definition is that the posterior measure is absolutely continuous with respect to μ_0 and measurable with respect to Σ. The interpretation of $\tilde{\mu}^N$ is that the data will only affect the projection, $P^p\,\mathrm{X}$. We can now show that also under this approximation, we get convergence to the true posterior.

Theorem A.4. *Let the measure $\tilde{\mu}^N$ be defined by (7), and let the measure μ be defined by (6). If μ_0 satisfies Assumption A.1, then $d_{TV}(\mu,\tilde{\mu}^N)\to 0$ as $N\to\infty$ and $p\to\infty$.*

The proof is given in Appendix B.

B Proofs

Proof of Proposition 2.1. For the first moment, note that

$$\begin{aligned}
\mathbb{E}\left[\lambda(\boldsymbol{s})\right] &= \mathbb{E}\left[\exp(\mathrm{X}(\boldsymbol{s}))\right] \\
&= \sum_{k=1}^{K}\mathbb{E}\left[\exp(\mathrm{X}_k(\boldsymbol{s})+\mu_k(\boldsymbol{s}))|\,\mathrm{X}_0(\boldsymbol{s})+\mu_0(\boldsymbol{s})\in(c_{k-1},c_k]\right]\mathbb{P}\left(\mathrm{X}_0(\boldsymbol{s})+\mu_0(\boldsymbol{s})\in(c_{k-1},c_k]\right) \\
&= \sum_{k=1}^{K}\mathbb{E}\left[\exp(\mathrm{X}_k(\boldsymbol{s})+\mu_k(\boldsymbol{s}))\right]\mathbb{P}\left(\mathrm{X}_0(\boldsymbol{s})+\mu_0(\boldsymbol{s})\in(c_{k-1},c_k]\right) \\
&= \sum_{k=1}^{K}\exp\left(\mu_k(\boldsymbol{s})+\frac{\mathcal{C}_k(0)}{2}\right)\mathbb{P}\left(\mathrm{X}_0(\boldsymbol{s})+\mu_0(\boldsymbol{s})\in(c_{k-1},c_k]\right),
\end{aligned}$$

where the final equality follows from the explicit form of the expectation of a log-normal random variable.

The second moment follows by similar calculations when considering that both points in space can be part of one of the K classes. Using the covariance of a bivariate log-normal distribution and defining

$$\begin{aligned}
p_{lk} &:= \mathbb{P}\left(\mathrm{X}_0(\boldsymbol{s}_1)+\mu_0(\boldsymbol{s}_1)\in(\mathrm{c}_{k-1},\mathrm{c}_k]\cap\mathrm{X}_0(\boldsymbol{s}_2)+\mu_0(\boldsymbol{s}_2)\in(\mathrm{c}_{l-1},\mathrm{c}_l]\right) \\
&= \int_{\mathrm{c}_{k-1}}^{\mathrm{c}_k}\left(\Phi\left(\frac{c_l-\mu^*(u)}{\sigma^*(u)}\right)-\Phi\left(\frac{c_{l-1}-\mu^*(u)}{\sigma^*(u)}\right)\right)\frac{e^{-\frac{(u-\mu_0(\boldsymbol{s}_1))^2}{2\,\mathcal{C}_0(0)}}}{\sqrt{2\pi}}du,
\end{aligned}$$

where $\mu^*(u)=\mu_0(\boldsymbol{s}_2)+\frac{\mathcal{C}_0(|\,\boldsymbol{s}_1-\boldsymbol{s}_2\,|)}{\mathcal{C}_0(0)}(u-\mu_0(\boldsymbol{s}_1))$ is the conditional expectation of $\mathrm{X}_0(\boldsymbol{s}_2)+\mu_0(\boldsymbol{s}_2)|\,\mathrm{X}_0(\boldsymbol{s}_1)+\mu_0(\boldsymbol{s}_1)=u$ and $\sigma^*(u)=\sqrt{\mathcal{C}_0(0)-\frac{\mathcal{C}_0(|\,\boldsymbol{s}_1-\boldsymbol{s}_2\,|)^2}{\mathcal{C}_0(0)}}$ is the corresponding conditional standard deviation.

Remember that $\mathcal{C}_0(0) = 1$ to make the model identifiable with respect to the threshold parameters. p_{lk} can then be defined through

$$\begin{aligned}
&\int_{-\infty}^{\mathrm{c}_k} \Phi\left(\frac{c_l - \mu^*(u)}{\sigma^*(u)}\right) \frac{e^{-\frac{(u-\mu_0(\boldsymbol{s}_1))^2}{2}}}{\sqrt{2\pi}} du \\
&\quad = \int_{-\infty}^{\mathrm{c}_k} \Phi\left(\frac{\mathrm{c}_l - \mu_0(\boldsymbol{s}_2)}{\sigma^*(u)} + \mu_0(\boldsymbol{s}_1)\frac{\mathcal{C}_0(|\,\boldsymbol{s}_1 - \boldsymbol{s}_2\,|)}{\sigma^*(u)} - u\frac{\mathcal{C}_0(|\,\boldsymbol{s}_1 - \boldsymbol{s}_2\,|)}{\sigma^*(u)}\right) \psi\left(u - \mu_0(\boldsymbol{s}_1)\right) du \\
&\quad = \int_{-\infty}^{\mathrm{c}_k} \Phi\left(a + ub\right) \psi\left(u - \mu_0(\boldsymbol{s}_1)\right) du = \int_{-\infty}^{\mathrm{c}_k - \mu_0(\boldsymbol{s}_1)} \Phi\left(a + b\mu_0(\boldsymbol{s}_1) + yb\right) \psi\left(y\right) dy,
\end{aligned}$$

where the variable substitution, $y = u - \mu_0(\boldsymbol{s}_1)$, was used. Here, $a = \frac{\mathrm{c}_l - \mu_0(\boldsymbol{s}_2)}{\sigma^*(u)} + \mu_0(\boldsymbol{s}_1)\frac{\mathcal{C}_0(|\,\boldsymbol{s}_1 - \boldsymbol{s}_2\,|)}{\sigma^*(u)}$, and $b = -\frac{\mathcal{C}_0(|\,\boldsymbol{s}_1 - \boldsymbol{s}_2\,|)}{\sigma^*(u)}$. Using $10,010.1$ from [43] then yields

$$\int_{-\infty}^{\mathrm{c}_k - \mu_0(\boldsymbol{s}_1)} \Phi\left(a + b\mu_0(\boldsymbol{s}_1) + yb\right) \psi\left(y\right) dy = \frac{1}{2\pi\sqrt{1-d^2}} \int_{\infty}^{\mathrm{c}_k - \mu_0(\boldsymbol{s}_1)} \int_{\infty}^{\frac{a+b\mu_0(\boldsymbol{s}_1)}{\sqrt{1+b^2}}} e^{-\frac{x^2+y^2-2xyd}{2(1-d^2)}} dxdy,$$

where $d = \frac{-b}{\sqrt{1+b^2}}$. Finally, $1 + b^2 = 1 + \frac{\mathcal{C}_0(|\,\boldsymbol{s}_1 - \boldsymbol{s}_2\,|)^2}{\sigma^*(u)^2} = \frac{1}{\sigma^*(u)^2}$ and hence $\frac{a+b\mu_0(\boldsymbol{s})}{\sqrt{1+b^2}} = c_l - \mu_0(\boldsymbol{s}_2)$ and $d = \mathcal{C}_0(|\,\boldsymbol{s}_1 - \boldsymbol{s}_2\,|)$.

This last integral corresponds to evaluating a cumulative distribution function of a bivariate centered Gaussian distribution with correlation d and unit marginal variances.

□

Proof of Proposition 2.2. The inhomogeneous empty space function, $F(\boldsymbol{s}_0, r)$ is defined as the probability of having at least one point inside a ball of radius r centered at $\boldsymbol{s}_0$, i.e. $F(\boldsymbol{s}_0, r) = \mathbb{P}\left(N(\mathrm{Y}; B(\boldsymbol{s}_0, r)) > 0\right)$. Here, $N(\mathrm{Y}; A)$ is the number of points inside the domain A for a realization of the point process, Y. Hence $F(\boldsymbol{s}_0, r) = 1 - \mathbb{P}\left(N(\mathrm{Y}; B(\boldsymbol{s}_0, r)) = 0\right)$. Now,

$$\begin{aligned}
\mathbb{P}\left(N(B(\mathrm{Y}; \boldsymbol{s}_0, r)) = 0\right) &= \mathbb{E}\left[\exp\left(-\int_{B(\boldsymbol{s}_0,r)} e^{\sum_{k=1}^K \mathrm{Z}_k(\boldsymbol{s})(\mathrm{X}_k(\boldsymbol{s}) + \mu_k(\boldsymbol{s}))} d\,\boldsymbol{s}\right)\right] \\
&= \mathbb{E}\left[\exp\left(-\int_{B(\boldsymbol{s}_0,r)} \sum_{k=1}^K \mathrm{Z}_k(\boldsymbol{s}) e^{\mathrm{X}_k(\boldsymbol{s}) + \mu_k(\boldsymbol{s})} d\,\boldsymbol{s}\right)\right] \\
&= \mathbb{E}\left[\prod_{k=1}^K \exp\left(-\int_{\mathcal{D}_k \cap B(\boldsymbol{s}_0,r)} e^{\mu_k(\boldsymbol{s})} e^{\mathrm{X}_k(\boldsymbol{s})} d\,\boldsymbol{s}\right)\right].
\end{aligned}$$

□

Due to the product space interpretation of X as the collection $\{\mathrm{X}_k\}_k$, we define norms on X as $\|\,\mathrm{X}\,\|_{(\cdot)} = \sum_{k=0}^K \|\,\mathrm{X}_k\,\|_{(\cdot)}$. That is, a norm on realizations of all Gaussian random fields jointly are defined as the sum of the norm for each of the $K + 1$ fields.

To simplify the proofs we note that the potential Φ can be written as a composition of two functions: The potential $\Phi(\mathrm{X}; \mathrm{Y}) = \Phi_P(G(\mathrm{X}); \mathrm{Y})$ where $\Phi_P : L_2(\mathcal{D}) \times \mathcal{Y} \to \mathbb{R}$ is the continuous Poisson log-likelihood function and $G : \mathcal{H}^\alpha \to L_2(\mathcal{D})$ is

$$G(X) = \sum_{k=1}^K \pi_k(\cdot) X_k(\cdot) = \log(\lambda(\cdot)),$$

where π_k is the classification function, $\pi_k(\boldsymbol{s}) = \mathbb{I}\left(\mathrm{c}_{k-1} \leq \mathrm{X}_0(\boldsymbol{s}) < \mathrm{c}_k\right)$. Similarly $\Phi^N(\mathrm{X};\mathrm{Y}) = \Phi_P^N(G(\mathrm{X});\mathrm{Y})$ where Φ_P^N is the Poisson log-likelihood function for the discretized domain.

To prove Proposition A.2, we will need two lemmas, where the first gives bounds for the likelihood potentials.

Lemma B.1. *Let $\|\,\mathrm{Y}\,\|_{\mathcal{Y}}$ denote the number of points in a given point pattern. For Φ in (4) and Φ^N in (5) we then have that:*

(i) For every $r > 0$, $\epsilon > 0$, and $s > 1$ with $\mathrm{X} \in \mathcal{H}^s$ and $\mathrm{Y} \in \mathcal{Y}$ with $||\,\mathrm{Y}\,||_{\mathcal{Y}} \leq r$, there exists a constant $M(\epsilon, r) \in \mathbb{R}$ such that $\Phi(\mathrm{X};\mathrm{Y}) \geq M(\epsilon, r) - \epsilon||\,\mathrm{X}\,||_s^2$.

(ii) For every $r > 0$, and $s > 1$ all $\mathrm{X} \in \mathcal{H}^s$ and all $\mathrm{Y} \in \mathcal{Y}$ with $\max\{||\,\mathrm{X}\,||_s, ||\,\mathrm{Y}\,||_{\mathcal{Y}}\} < r$ we have $\Phi(\mathrm{X};\mathrm{Y}) \leq |\,\mathcal{D}\,|e^{Cr} + C^2r^2$.

Proof. To show (i) note that

$$\begin{aligned}\Phi_P(G(\mathrm{X});\mathrm{Y}) &= \int_{\mathcal{D}} \exp\left(G(\mathrm{X})\right) ds - \sum_{s_j \in \mathrm{Y}} G(\mathrm{X}) \geq -\sum_{s_j \in \mathrm{Y}} G(\mathrm{X}) \geq -\|\,\mathrm{Y}\,\|_{\mathcal{Y}} \|G(\mathrm{X})\|_{L^\infty(\mathcal{D})} \\ &\geq -r\|G(\mathrm{X})\|_{L^\infty(\mathcal{D})}.\end{aligned}$$

By Assumption A.1 and the Sobolev embedding theorem we have that $\|\,\mathrm{X}\,\|_{L^\infty(\mathcal{D})} \leq C\|\,\mathrm{X}\,\|_s$. Thus $\|G(\mathrm{X})\|_{L^\infty(\mathcal{D})} \leq \|\,\mathrm{X}\,\|_{L^\infty(\mathcal{D})} \leq C\|\,\mathrm{X}\,\|_s$ and we have $\Phi_P(G(\mathrm{X}), \mathrm{Y}) \geq -rC\|\,\mathrm{X}\,\|_s$. Now, $0 \leq (\frac{Cr}{2\sqrt{\epsilon}} - \sqrt{\epsilon}\|\,\mathrm{X}\,\|_s)^2 = \frac{C^2r^2}{4\epsilon} + \epsilon\|\,\mathrm{X}\,\|_s^2 - Cr\|\,\mathrm{X}\,\|_s$. Hence

$$Cr\|\,\mathrm{X}\,\|_s \leq \epsilon\|\,\mathrm{X}\,\|_s^2 + \frac{C^2r^2}{4\epsilon} = \epsilon\|\,\mathrm{X}\,\|_s^2 - M(\epsilon, r).$$

By the same argument,

$$\Phi_P^N(G(\mathrm{X});\mathrm{Y}) = \sum_{i \in I^N} \left(|\,\mathcal{D}_i\,| \exp\left(G(\mathrm{X})(s)\right) - \mathrm{Y}_i\, G(\mathrm{X})(s_i)\right) \geq M(\epsilon, r) - \epsilon\|\,\mathrm{X}\,\|_s^2.$$

Statement (ii) holds for Φ since

$$\begin{aligned}\Phi_P(G(\mathrm{X});\mathrm{Y}) &= \int_{\mathcal{D}} \exp\left(G(\mathrm{X})(\boldsymbol{s})\right) d\,\boldsymbol{s} - \sum_{\boldsymbol{s}_j \in \mathrm{Y}} G(\mathrm{X})(\boldsymbol{s}_j) \\ &\leq |\,\mathcal{D}\,|e^{\|G(\mathrm{X})\|_{L^\infty(\mathcal{D})}} + \|\,\mathrm{Y}\,\|_{\mathcal{Y}}\|G(\mathrm{X})\|_{L^\infty(\mathcal{D})} \\ &\leq |\,\mathcal{D}\,|e^{Cr} + Cr^2 \leq |\,\mathcal{D}\,|e^{Cr} + C^2r^2,\end{aligned}$$

and the same for Φ^N since

$$\Phi_P^N(G(\mathrm{X}), \mathrm{Y}) = \sum_{i \in I^N} \left(|\,\mathcal{D}_i\,|e^{\|G(\mathrm{X})\|_s} - \mathrm{Y}_i\, G(\mathrm{X})(s_i)\right) \leq |\,\mathcal{D}\,|e^{Cr} + C^2r^2.$$

□

The second lemma we need concerns the regularity of the level sets of X_0. Let $S_k^0(\mathrm{X}_0) = \{\boldsymbol{s} : \mathrm{X}_0(\boldsymbol{s}) = \mathrm{c}_k\}$ be the level set of X_0 for the level c_k and set $S^0(\mathrm{X}_0) = \cup_{k=1}^K S_k^0(\mathrm{X}_0)$. Further, let J_s denote the set of indices for all subregions $\mathcal{D}_j$ that do not intersect with $S^0(\mathrm{X}_0)$, that is, $j \in J_s$ if $\mathcal{D}_j \cap \mathcal{D}_k^0 = \emptyset$ for all $1 \leq k \leq K$, and define $S(\mathrm{X}) = \cup_{j \in J_s} \mathcal{D}_j$ as the set of all subregions where the level sets are not included. We then have the following result about $S(\mathrm{X})$, and $\mathcal{L}_d(S(\mathrm{X}))$where $\mathcal{L}_d$ denotes the Lebesgue measure in dimension d.

Lemma B.2. *Let Assumption A.1 hold, then*

- $\mathcal{L}_2(S^0(\mathrm{X})) = 0$ *a.s.*
- $\mathbb{E}\left[\mathcal{L}_2(S^C(\mathrm{X}))\right] \to 0$ *as* $N \to \infty$.
- *For any finite set of points* Y, $\mathbb{E}\left[\|S^C(\mathrm{X}) \cap \mathrm{Y}\,\|_{\mathcal{Y}}\right] \to 0$ *as* $N \to \infty$.

Proof. That $\mathcal{L}_2(S^0(\mathrm{X})) = 0$ a.s. follows from Proposition 2.8 in Iglesias et al. [29].

We will now show that $\mathbb{E}\left[\mathcal{L}_2(S^C(\mathrm{X}))\right]$ goes to zero. Note that a curve segment of length l can at most cover $4(\frac{l}{h}+1)$ subregions $\mathcal{D}_j$. Hence, the number of subregions $\mathcal{D}_j$ that have a level crossing, $N - |J_s|$, is bounded by $\sum_{i=1}^{N^*} 4(l_i/h+1)$, where N^* is the number of disjoint line segments in $S^0(X_0)$ and l_i the length of ith segment. This gives that

$$\mathbb{E}\left[\mathcal{L}_2(S^C(\mathrm{X}))\right] \leq h^2\left(\frac{4}{h}\mathbb{E}\left[\mathcal{L}_1(S^0(X_0))\right] + 4\mathbb{E}\left[N^*\right]\right) \leq 4h(\mathbb{E}\left[\mathcal{L}_1(S^0(X_0))\right] + hN^*).$$

Thus, the result follows if we can bound $\mathbb{E}\left[\mathcal{L}(S^0(X_0)))\right]$ and $\mathbb{E}\left[N^*\right]$. By assumption X_0 satisfies the conditions of Rice Theorem [3], which gives that $\mathbb{E}\left[\mathcal{L}_1(S^0(X_0)))\right] < \infty$. Let N_k denote the number of local maxima of X_0 over the level c_k and let $N^0 = \sum_k N_k$. Since $\mathbb{E}\left[N^*\right]$ is bounded by $\mathbb{E}\left[N^0\right]$, and Rice Theorem bounds $\mathbb{E}\left[N^0\right]$, the result follows.

Finally, we show that $\mathbb{E}\left[\|S^C(\mathrm{X}) \cap \mathrm{Y}\,\|_{\mathcal{Y}}\right]$ goes to zero. We only consider the case $K = 1$ and $Y = \{y\}$, as the general result follows directly given that the claim holds for this special case. Let $B(y, \epsilon_N)$ be a ball centered at y, where ϵ_N is chosen so that the subregions covering y are contained in the ball. To prove the result we need to show that $\mathbb{P}(\mathcal{L}_i(B(y,\epsilon_N) \cap X_0^{-1}(c_1)) > 0) \to 0$ as $N \to 0$, for both $i = 0, 1$, where $\mathcal{L}_i$ are the Lipschitz-Killing curvatures. Since X_0 is a Morse function and $\mathcal{L}_i(B(y,\epsilon_N)) \to 0$, Theorem 15.9.4 in [2] shows that $\mathbb{E}[\mathcal{L}_i(B(y,\epsilon_N) \cap X_0^{-1}(c_1)))] \to 0$ for $i = 0, 1$. Thus $\mathbb{P}(\mathcal{L}_1(B(y,\epsilon_N) \cap X_0^{-1}(c_1)) > 0) \to 0$ as $\mathcal{L}_1(B(y,\epsilon_N) \cap X_0^{-1}(c_1))$ is non-negative random variable. Since any $B(y,\epsilon_N)$ converges to a point, it follows that $\mathcal{L}_0(B(y,\epsilon_N) \cap X_0^{-1}(c_1))$ (the Euler characteristic) converges to a non-negative random variable, and thus $\mathbb{P}(\mathcal{L}_0(B(y,\epsilon_N) \cap X_0^{-1}(c_1)) > 0) \to 0$.

□

Proof of Proposition A.2. We only state the proof for μ since the proof for μ^N follows similarly. To show the result we must show that the Φ is a measurable function, and then that the measure is normalizable. To prove measurability it suffices, by Lemma 6.1 in Iglesias et al. [29], to show that that Φ is continuous μ_0-almost surely. Thus for $\hat{X}, \tilde{X} \in \mathcal{H}^s, s > 1$, we must show that $|\,\Phi(\hat{\mathrm{X}}) - \Phi(\tilde{\mathrm{X}})| \to 0$ as $\|\hat{\mathrm{X}} - \tilde{\mathrm{X}}\|_s \to 0$. Note that

$$|\,\Phi(\hat{\mathrm{X}}) - \Phi(\tilde{\mathrm{X}})| \leq \int_{\mathcal{D}} |e^{G(\hat{\mathrm{X}})(\boldsymbol{s})} - e^{G(\tilde{\mathrm{X}})(\boldsymbol{s})}| d\,\boldsymbol{s} + \sum_{\boldsymbol{s}_j \in \mathrm{Y}} |G(\hat{\mathrm{X}})(\boldsymbol{s}_j) - G(\tilde{\mathrm{X}})(\boldsymbol{s}_j)|. \tag{8}$$

We show continuity of the two terms separately. For the first term in (8) it follows that

$$\begin{aligned}\int_{\mathcal{D}} |e^{G(\hat{\mathrm{X}})(\boldsymbol{s})} - e^{G(\tilde{\mathrm{X}})(\boldsymbol{s})}| d\,\boldsymbol{s} &\leq \int_{\mathcal{D}} \exp\left(|G(\hat{\mathrm{X}})(\boldsymbol{s})| + |G(\tilde{\mathrm{X}})(\boldsymbol{s})|\right) |G(\hat{\mathrm{X}})(\boldsymbol{s}) - G(\tilde{\mathrm{X}})(\boldsymbol{s})| d\,\boldsymbol{s} \\ &\leq C|\,\mathcal{D}\,| e^{C\|\hat{\mathrm{X}}\|_s + C\|\tilde{\mathrm{X}}\|_s} \|G(\hat{\mathrm{X}}) - G(\tilde{\mathrm{X}})\|_s.\end{aligned}$$

Here the first inequality is due to the mean value theorem, and the second inequality comes from using Sobolev's embedding theorem, and Hölders inequality. Since $\|G(\hat{\mathrm{X}}) - G(\tilde{\mathrm{X}})\|_s \leq \sum_{k=1}^K \|\,\mathrm{X}\,\|_s \|\pi_k(\hat{\mathrm{X}}_0) - \pi_k(\tilde{\mathrm{X}}_0)\|_s + \|\hat{\mathrm{X}} - \tilde{\mathrm{X}}\|_s$, it suffices to show that π_k is continuous. By Lemma B.2, $\mathcal{L}(S^0(\mathrm{X})) = 0$ a.s. and since $\pi_k(\cdot)$ is constant on $S^0(\mathrm{X})^C$ it is also a.s. continuous. By

Proposition 2.6 in Iglesias et al. [29], $\pi_k(\cdot)$ is therefore continuous on $L_2(\mathcal{D})$ and thus also on $\mathcal{H}^1$ since it is a.s. constant.

The second term in (8) can be bounded by $C\|\,\mathrm{Y}\,\|_{\mathcal{Y}}\|\hat{\mathrm{X}}-\tilde{\mathrm{X}}\|_s$ a.s. since $|G(\hat{\mathrm{X}})(\boldsymbol{s})-G(\tilde{\mathrm{X}})(\boldsymbol{s})| \leq \|\hat{\mathrm{X}}-\tilde{\mathrm{X}}\|_{L^\infty(\mathcal{D})}$. Finally, by Lemma B.1 the function Φ is bounded from above and below, and thus the measure can be normalized. □

From here on we will simplify the notation by omitting the observed point pattern from the likelihood potential and the constants, i.e. $\Phi(\mathrm{X}) = \Phi(\mathrm{X};\mathrm{Y})$ and $C_\mu = C_\mu(\mathrm{Y})$.

Proof of Theorem A.3. By Stuart [56, Lemma 6.36], the Hellinger distance bounds the total variation norm, so it suffices to show convergence in Hellinger distance. Take $X \in \mathcal{H}^s, s > 1$. By the triangle inequality,

$$
\begin{aligned}
2d_{\mathrm{Hell}}(\mu,\mu^N)^2 &= \int\left(\sqrt{\frac{d\mu}{d\nu}}-\sqrt{\frac{d\mu^N}{d\nu}}\right)^2 d\mu_0(\mathrm{X}) = \int\left(\frac{e^{-\frac{1}{2}\,\Phi(\mathrm{X})}}{\sqrt{C_\mu}}-\frac{e^{-\frac{1}{2}\,\Phi^N(\mathrm{X})}}{\sqrt{C_{\mu^N}}}\right)^2 d\mu_0(\mathrm{X}) \\
&\leq \frac{1}{C_\mu}\int\left|e^{-\frac{1}{2}\,\Phi(\mathrm{X})}-e^{-\frac{1}{2}\,\Phi^N(\mathrm{X})}\right|^2 d\mu_0(\mathrm{X}) + \left|\frac{1}{\sqrt{C_\mu}}-\frac{1}{\sqrt{C_{\mu^N}}}\right|^2\int e^{-\,\Phi^N(\mathrm{X})}d\mu_0(\mathrm{X}) \\
&= I_1 + I_2,
\end{aligned}
$$

where $C_\mu = \int e^{-\,\Phi(\mathrm{X})}d\mu_0(\mathrm{X})$ and $C_{\mu^N} = \int e^{-\,\Phi^N(\mathrm{X})}d\mu_0(\mathrm{X})$. We now first show that I_2 can be bounded by I_1 and then show that $I_1 \to 0$ as $N \to \infty$. Note that

$$
\begin{aligned}
I_2 &\leq (C_\mu - C_{\mu^N})^2\frac{1}{4}\left(\min\{C_\mu, C_{\mu^N}\}\right)^{-3} C_{\mu^N} \\
&= \left|\int e^{-\,\Phi(\mathrm{X})}-e^{-\,\Phi^N(\mathrm{X})}d\mu_0(\mathrm{X})\right|^2\frac{1}{4}\left(\min\{C_\mu, C_{\mu^N}\}\right)^{-3} C_{\mu^N} \\
&\leq \frac{C_{\mu^N}}{4\min\{C_\mu, C_{\mu^N}\}^3}\left(\int\left|e^{-\,\Phi(\mathrm{X})}-e^{-\,\Phi^N(\mathrm{X})}\right|d\mu_0(\mathrm{X})\right)^2 \\
&\leq \frac{C_{\mu^N}}{4\min\{C_\mu, C_{\mu^N}\}^3}\int\left|e^{-\frac{1}{2}\,\Phi(\mathrm{X})}-e^{-\frac{1}{2}\,\Phi^N(\mathrm{X})}\right|^2 d\mu_0(\mathrm{X})\int e^{\epsilon\|\,\mathrm{X}\,\|_1^2 - M(\epsilon,\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}d\mu_0(\mathrm{X}) \\
&\leq CI_1.
\end{aligned}
$$

Here the third inequality is due to Hölder's inequality and Ferniques theorem [15, Theorem A.3]. Now to bound I_1 note that

$$
I_1 \leq \frac{1}{4C_\mu}\int e^{\epsilon\|\,\mathrm{X}\,\|_s^2 - M(\epsilon,\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}|\,\Phi(\mathrm{X}) - \Phi^N(\mathrm{X})|^2 d\mu_0(\mathrm{X}).
$$

Since the function G is Lipschitz continuous on $S(X)$ (see Lemma B.2) we get

$$
\begin{aligned}
\left|\Phi(\mathrm{X}) - \Phi^N(\mathrm{X})\right| \leq\; & Ce^{C\|\,\mathrm{X}\,\|_s}|\,\mathcal{D}\,|h + C\|\,\mathrm{Y}\,\|_{\mathcal{Y}}h \\
& + C\|\,\mathrm{X}\,\|_s(e^{C\|\,\mathrm{X}\,\|_s}\mathcal{L}(S^C(\mathrm{X})) + \|S^C(\mathrm{X})\cap\mathrm{Y}\,\|_{\mathcal{Y}}),
\end{aligned}
$$

and thus

$$
\begin{aligned}
I_1 \leq\; & \frac{2}{4C_\mu}C\int e^{3\epsilon\|\,\mathrm{X}\,\|_s^2 - 3M(\epsilon,1+\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}(|\,\mathcal{D}\,| + \|\,\mathrm{Y}\,\|_{\mathcal{Y}})^2h^2 d\mu_0(\mathrm{X}) \\
& + \frac{2}{4C_\mu}C\int e^{3\epsilon\|\,\mathrm{X}\,\|_s^2 - 3M(\epsilon,1+\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}(\mathcal{L}(S^C(\mathrm{X})) + \|S^C(\mathrm{X})\cap\mathrm{Y}\,\|_{\mathcal{Y}})^2 d\mu_0(\mathrm{X}).
\end{aligned}
$$

Now the first integral on the right hand side clearly goes to zero as $N \to \infty$. The second integral can be bounded by

$$\begin{aligned}
&\sqrt{\mathbb{E}\left[e^{6\epsilon\|\,\mathrm{X}\,\|_s^2 - 6M(\epsilon,\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}\right]}\sqrt{\mathbb{E}\left[(\mathcal{L}(S^C(\mathrm{X})) + \|S^C(\mathrm{X}) \cap \mathrm{Y}\,\|_{\mathcal{Y}})^4\right]} \\
&\leq C_2\sqrt{\mathbb{E}\left[(\mathcal{L}(S^C(\mathrm{X})) + \|S^C(\mathrm{X}) \cap \mathrm{Y}\,\|_{\mathcal{Y}})^4\right]} \\
&\leq C_2(\mathcal{L}(\mathcal{D}) + \|\,\mathrm{Y}\,\|_{\mathcal{Y}})^3\left(\mathbb{E}\left[\mathcal{L}(S^C(\mathrm{X}))\right] + \mathbb{E}\left[\|S^C(\mathrm{X}) \cap \mathrm{Y}\,\|_{\mathcal{Y}}\right]\right),
\end{aligned}$$

and as $N \to \infty$ this also goes to zero by Lemma B.2. $\square$

Proof of Theorem A.4. Denote the posterior measure for the fully discretized model by $\tilde{\mu}^N$. The TV distance between the posterior measures can be bounded as

$$d_{TV}(\mu, \tilde{\mu}^N) \leq d_{TV}(\mu, \mu^N) + d_{TV}(\mu^N, \tilde{\mu}^N),$$

where the first term goes to zero by theorem A.3. Clearly, $\tilde{\mu}^N$ as given in (7) defines a posterior measure with respect to μ_0 by the same arguments as in the proof of Proposition A.2, and it coincides with μ^N on the span of $\{e_j\}_{j>p+1}$. We can therefore bound $d_{TV}(\mu^N, \tilde{\mu}^N)$ using the same method as in the proof of theorem A.3, this gives that $2d_{\mathrm{Hell}}(\mu^N, \tilde{\mu}^N)^2 \leq I_1 + I_2$, where now,

$$\begin{aligned}
I_1 &= \frac{1}{C_{\mu^N}}\int_{\mathcal{X}}\left|e^{-\frac{1}{2}\,\Phi^N(\mathrm{X})} - e^{-\frac{1}{2}\,\Phi^N(P^p\,\mathrm{X})}\right|^2 d\mu_0(\mathrm{X}) \\
I_2 &= \left|\frac{1}{\sqrt{C_{\mu^N}}} - \frac{1}{\sqrt{C_{\tilde{\mu}^N}}}\right|^2 \int e^{-\,\Phi^N(P^p\,\mathrm{X})} d\mu_0(\mathrm{X}).
\end{aligned}$$

We can again bound I_2 by CI_1, so what remains to be shown is that I_1 goes to zero as $p \to \infty$. Let $X \in \mathcal{H}^s, s > 1$. Since P^p is a projection, we then clearly have that $\|P^pX\|_s \leq \|X\|_s$. By Lemma B.1(i) and Hölders inequality

$$\begin{aligned}
I_1 &\leq \frac{1}{4C_\mu}\int e^{\epsilon\|\,\mathrm{X}\,\|_s^2 - M(\epsilon,\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}|\,\Phi^N(\mathrm{X}) - \Phi^N(P^p\,\mathrm{X})|d\mu_0(\mathrm{X}) \\
&\leq C\sqrt{\int e^{\epsilon\|\,\mathrm{X}\,\|_s^2 - M(\epsilon,\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}d\mu_0(\mathrm{X})\mathbb{E}\left[|\,\Phi^N(\mathrm{X}) - \Phi^N(P^p\,\mathrm{X})|^2\right]}.
\end{aligned}$$

We will now focus on bounding the expectation above. Using Ferniques theorem

$$\begin{aligned}
\left|\Phi^N(\mathrm{X}) - \Phi^N(P^p\,\mathrm{X})\right| &= \sum_{i=1}^N |\,\mathcal{D}_i\,|(e^{G(\mathrm{X})(\boldsymbol{s}_i)} - e^{G(P^p\,\mathrm{X})(\boldsymbol{s}_i)}) - \mathrm{Y}_i(G(P^p\,\mathrm{X})(\boldsymbol{s}_i) - G(\mathrm{X})(\boldsymbol{s}_i)) \\
&\leq e^{\epsilon\|\,\mathrm{X}\,\|_s^2 - M(\epsilon,\|\,\mathrm{Y}\,\|_{\mathcal{Y}})}\sum_{i=1}^N(|\,\mathcal{D}_i\,| + Y_i)|G(P^p\,\mathrm{X})(\boldsymbol{s}_i) - G(\mathrm{X})(\boldsymbol{s}_i)|.
\end{aligned}$$

Using the inequalities

$$\begin{aligned}
|G(P^p\,\mathrm{X})(\boldsymbol{s}) - G(\mathrm{X})(\boldsymbol{s})| &\leq \sum_{k=1}^K |\pi_k(\mathrm{X}_0)(\boldsymbol{s})\,\mathrm{X}_k(\boldsymbol{s}) - \pi_k(P^p\,\mathrm{X}_0)(\boldsymbol{s})P^p\,\mathrm{X}_k(\boldsymbol{s})| \\
&\leq \sum_{k=1}^K (|\mathrm{X}_k(\boldsymbol{s}) - P^p\,\mathrm{X}_k(\boldsymbol{s})| + C\|\,\mathrm{X}\,\|_s\,|\pi_k(\mathrm{X}_0)(\boldsymbol{s}) - \pi_k(P^p\,\mathrm{X}_0)(\boldsymbol{s})|)
\end{aligned}$$

yields that

$$\mathbb{E}\left[|\,\Phi^N(\mathrm{X}) - \Phi^N(P^p\,\mathrm{X})|^2\right] \leq C\mathbb{E}\left[\left(\sum_{i\in I^N}(|\,\mathcal{D}_i\,| + Y_i)\sum_{k=1}^{K}|\mathrm{X}_k(\boldsymbol{s}_i) - P^p\,\mathrm{X}_k(\boldsymbol{s}_i)|\right)^2\right] + C\mathbb{E}\left[\left(\sum_{i\in I^N}(|\,\mathcal{D}_i\,| + Y_i)\sum_{k=1}^{K}|\pi_k(\mathrm{X}_0)(\boldsymbol{s}_i) - \pi_k(P^p\,\mathrm{X}_0)(\boldsymbol{s}_i)|\right)^2\right]. \tag{9}$$

Note that $|D_i| \propto N^{-1}$ and that $X(\boldsymbol{s})$ is bounded for each $\boldsymbol{s} \in \mathcal{D}$ almost surely. Let $Q^pX = X - P^pX$ and note that Q^pX for each $\boldsymbol{s} \in \mathcal{D}$ is a mean-zero Gaussian variable with a variance σ_p^2 that goes to zero as $p \to \infty$. Thus, the first term in (9) clearly goes to zero as $p \to \infty$. Since $|\pi_k(\mathrm{X}_0)(\boldsymbol{s}_i) - \pi_k(P^p\,\mathrm{X}_0)(\boldsymbol{s}_i)|$ is bounded by one, the second term in (9) can be bounded by

$$C\sum_{i=1}^{N}(|\,\mathcal{D}_i\,| + Y_i)\mathbb{E}\left[\sum_{k=1}^{K}|\pi_k(\mathrm{X}_0)(\boldsymbol{s}_i) - \pi_k(P^p\,\mathrm{X}_0)(\boldsymbol{s}_i)|\right].$$

Here the expectation can be bounded as

$$\mathbb{E}\left[\sum_{k=1}^{K}|\pi_k(\mathrm{X}_0)(\boldsymbol{s}_i) - \pi_k(P^p\,\mathrm{X}_0)(\boldsymbol{s}_i)|\right] \leq K\max_k\{\,\mathbb{P}\left(\mathrm{X}_0(\boldsymbol{s}_i) \leq \mathrm{c}_k \cap P^p\,\mathrm{X}_0(\boldsymbol{s}_i) > \mathrm{c}_k\right) + \mathbb{P}\left(\mathrm{X}_0(\boldsymbol{s}) > \mathrm{c}_k \cap P^p\,\mathrm{X}_0(\boldsymbol{s}) < \mathrm{c}_k\right)\}.$$

We now show how to bound the first probability, and the second probability is bounded by similar calculations. Define the events $A = \{\mathrm{X}_0(\boldsymbol{s}_i) \leq \mathrm{c}_k \cap P^p\,\mathrm{X}_0(\boldsymbol{s}_i) > \mathrm{c}_k\}$ and $B = \{P^pX_0(s_i) \in [c_k, c_k + \epsilon]\}$. It follows that

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|B^C)\mathbb{P}(B^C) \leq \mathbb{P}(B) + \mathbb{P}(A|B^C)\\ &\leq \mathbb{P}(P^pX_0(s_i) \in [c_k, c_k + \epsilon]) + \mathbb{P}(Q^pX_0(s_i) \leq -\epsilon).\end{aligned}$$

Now set $\epsilon = \sqrt{\sigma_p}$ and recall that $\mathbb{P}\left(Z > t\right) < \frac{1}{\sqrt{2\pi}t}e^{-t^2/2}$ if $Z \sim \mathsf{N}(0,1)$. This gives that

$$\mathbb{P}\left(A\right) \leq \mathbb{P}\left(0 < P^p\,\mathrm{X}_0(\boldsymbol{s}_i) \leq \sqrt{\sigma_p}\right) + \mathbb{P}\left(Q^p\,\mathrm{X}_0(\boldsymbol{s}_i) \leq -\sqrt{\sigma_p}\right) \leq C\sqrt{\sigma_p} + \frac{\sqrt{\sigma_p}}{\sqrt{2\pi}}e^{-\frac{1}{2\sigma_p}},$$

which goes to zero as $p \to \infty$, and thus so does the final expectation in (9).

□